

MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women

Khan, Khalid S; Tryposkiadis, Konstantinos; Tirlapur, Seema A; Middleton, Lee J; Sutton, Andrew J; Priest, Lee; Ball, Elizabeth; Balogun, Moji; Sahdev, Anju; Roberts, Tracy; Birch, Judy; Daniels, Jane P; Deeks, Jonathan J

DOI:
[10.3310/hta22400](https://doi.org/10.3310/hta22400)

License:
None: All rights reserved

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Khan, KS, Tryposkiadis, K, Tirlapur, SA, Middleton, LJ, Sutton, AJ, Priest, L, Ball, E, Balogun, M, Sahdev, A, Roberts, T, Birch, J, Daniels, JP & Deeks, JJ 2018, 'MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women: a test-accuracy study and economic evaluation', *Health Technology Assessment*, vol. 22, no. 40, pp. 1-92. <https://doi.org/10.3310/hta22400>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:
Checked for eligibility: 02/07/2019

This document is the final Version of Record of a report that appeared in Health Technology Assessment published by NIHR Journals Library. It can be found at: <https://doi.org/10.3310/hta22400>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women: a test-accuracy study and economic evaluation

*Khalid S Khan, Konstantinos Tryposkiadis, Seema A Tirlapur, Lee J Middleton, Andrew J Sutton, Lee Priest, Elizabeth Ball, Moji Balogun, Anju Sahdev, Tracy Roberts, Judy Birch, Jane P Daniels and Jonathan J Deeks
on behalf of the MEDAL study group*

MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women: a test-accuracy study and economic evaluation

Khalid S Khan,¹ Konstantinos Tryposkiadis,²
Seema A Tirlapur,¹ Lee J Middleton,²
Andrew J Sutton,^{3,4} Lee Priest,² Elizabeth Ball,⁵
Moji Balogun,⁶ Anju Sahdev,⁷ Tracy Roberts,³
Judy Birch,⁸ Jane P Daniels^{2,9*}† and Jonathan J Deeks^{2†}
on behalf of the MEDAL study group

¹Women's Health Research Unit, Queen Mary University of London, London, UK

²Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

³Health Economics Unit, University of Birmingham, Birmingham, UK

⁴Leeds Institute of Health Sciences and NIHR Diagnostic Evidence Co-operative
Leeds, University of Leeds, Leeds, UK

⁵Obstetrics and Gynaecology, Barts Health NHS Trust, London, UK

⁶Radiology Department, Birmingham Women's and Children's NHS Foundation
Trust, Birmingham, UK

⁷Radiology Department, Barts Health NHS Trust, London, UK

⁸Pelvic Pain Support Network, Poole, UK

⁹Nottingham Clinical Trials Unit, University of Nottingham, Nottingham, UK

*Corresponding author

†Joint last author

Declared competing interests of authors: Elizabeth Ball declares UK travel reimbursement from Shire Medical (Lexington, MA, USA) outside the submitted work. Jonathan J Deeks was Deputy Chairperson of the National Institute for Health Research Health Technology Assessment (HTA) Commissioning Board (2011–16) and the HTA Efficient Study Designs Board (2016).

Published July 2018

DOI: 10.3310/hta22400

This report should be referenced as follows:

Khan KS, Tryposkiadis K, Tirlapur SA, Middleton LJ, Sutton AJ, Priest L, *et al.* MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women: a test-accuracy study and economic evaluation. *Health Technol Assess* 2018;**22**(40).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents*®/Clinical Medicine.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/22/50. The contractual start date was in October 2011. The draft report began editorial review in August 2016 and was accepted for publication in November 2017. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© Queen's Printer and Controller of HMSO 2018. This work was produced by Khan *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA and EME Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Director of the NIHR Dissemination Centre, University of Southampton, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Wellbeing Research, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

MRI versus laparoscopy to diagnose the main causes of chronic pelvic pain in women: a test-accuracy study and economic evaluation

Khalid S Khan,¹ Konstantinos Tryposkiadis,² Seema A Tirlapur,¹ Lee J Middleton,² Andrew J Sutton,^{3,4} Lee Priest,² Elizabeth Ball,⁵ Moji Balogun,⁶ Anju Sahdev,⁷ Tracy Roberts,³ Judy Birch,⁸ Jane P Daniels^{2,9*†} and Jonathan J Deeks^{2†} on behalf of the MEDAL study group

¹Women's Health Research Unit, Queen Mary University of London, London, UK

²Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

³Health Economics Unit, University of Birmingham, Birmingham, UK

⁴Leeds Institute of Health Sciences and NIHR Diagnostic Evidence Co-operative Leeds, University of Leeds, Leeds, UK

⁵Obstetrics and Gynaecology, Barts Health NHS Trust, London, UK

⁶Radiology Department, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK

⁷Radiology Department, Barts Health NHS Trust, London, UK

⁸Pelvic Pain Support Network, Poole, UK

⁹Nottingham Clinical Trials Unit, University of Nottingham, Nottingham, UK

*Corresponding author Jane.Daniels@nottingham.ac.uk

†Joint last author

Background: Chronic pelvic pain (CPP) symptoms in women are variable and non-specific; establishing a differential diagnosis can be hard. A diagnostic laparoscopy is often performed, although a prior magnetic resonance imaging (MRI) scan may be beneficial.

Objectives: To estimate the accuracy and added value of MRI in making diagnoses of (1) idiopathic CPP and (2) the main gynaecological causes of CPP. To quantify the impact MRI can have on decision-making with respect to triaging for therapeutic laparoscopy and to conduct an economic evaluation.

Design: Comparative test-accuracy study with cost-effectiveness modelling.

Setting: Twenty-six UK-based hospitals.

Participants: A total of 291 women with CPP.

Methods: Pre-index information concerning the patient's medical history, previous pelvic examinations and ultrasound scans was collected. Women reported symptoms and quality of life at baseline and 6 months. MRI scans and diagnostic laparoscopy (undertaken and interpreted blind to each other) were the index tests. For each potential cause of CPP, gynaecologists indicated their level of certainty that the condition was causing the pelvic pain. The analysis considered both diagnostic laparoscopy as a reference standard for observing structural gynaecological causes and consensus from a two-stage expert independent panel for ascertaining the cause of CPP. The stage 1 consensus was based on pre-index, laparoscopy and follow-up

data; for stage 2, the MRI scan report was also provided. The primary analysis involved calculations of sensitivity and specificity for the presence or absence of each structural gynaecological cause of pain. A decision-analytic model was developed, with a 6-month time horizon. Two strategies, laparoscopy or MRI, were considered and populated with study data.

Results: Using reference standards of laparoscopic and expert panel diagnoses, MRI scans had high specificity but poor sensitivity for observing deep-infiltrating endometriosis, endometrioma, adhesions and ovarian cysts. MRI scans correctly identified 56% [95% confidence interval (CI) 48% to 64%] of women judged to have idiopathic CPP, but missed 46% (95% CI 37% to 55%) of those considered to have a gynaecological structural cause of CPP. MRI added significant value, over and above the pre-index information, in identifying deep-infiltrating endometriosis ($p = 0.006$) and endometrioma ($p = 0.02$) as the cause of pain, but not for other gynaecological structural causes or for identifying idiopathic CPP ($p = 0.08$). Laparoscopy was significantly more accurate than MRI in diagnosing idiopathic CPP ($p < 0.0001$), superficial peritoneal endometriosis ($p < 0.0001$), deep-infiltrating endometriosis ($p < 0.0001$) and endometrioma of the ovary ($p = 0.02$) as the cause of pelvic pain. The accuracy of laparoscopy appeared to be able to rule in these diagnoses. Using MRI to identify women who require therapeutic laparoscopy would lead to 369 women in a cohort of 1000 receiving laparoscopy unnecessarily, and 136 women who required laparoscopy not receiving it. The economic analysis highlighted the importance of the time horizon, the prevalence of CPP and the cut-off values to inform the sensitivity and specificity of MRI and laparoscopy on the model results. MRI was not found to be a cost-effective diagnostic approach in any scenario.

Conclusions: MRI was dominated by laparoscopy in differential diagnosis of women presenting to gynaecology clinics with CPP. It did not add value to information already gained from history, examination and ultrasound about idiopathic CPP and various gynaecological conditions.

Trial registration: Current Controlled Trials ISRCTN13028601.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 22, No. 40. See the NIHR Journals Library website for further project information.

Contents

List of tables	xi
List of figures	xiii
List of abbreviations	xv
Plain English summary	xvii
Scientific summary	xix
Chapter 1 Introduction	1
Background	1
<i>Definition and prevalence of chronic pelvic pain</i>	1
<i>Personal and societal costs of chronic pelvic pain</i>	1
<i>Pathogenesis of chronic pelvic pain</i>	1
<i>Variation in the presentation of chronic pelvic pain</i>	2
Diagnosis of the cause of chronic pelvic pain	2
<i>Clinical assessment and ultrasound</i>	2
<i>Laparoscopy</i>	2
<i>The role of laparoscopy in the differential diagnosis of chronic pelvic pain</i>	3
<i>The accuracy of laparoscopy in the diagnosis of chronic pelvic pain</i>	3
<i>Magnetic resonance imaging in the diagnosis of chronic pelvic pain</i>	3
Treatment of chronic pelvic pain	4
Overview of the Magnetic resonance imaging for Establishing Diagnosis Against Laparoscopy study	5
Rationale for the Magnetic resonance imaging for Establishing Diagnosis Against Laparoscopy study	5
The choice of study design	7
Aim and objectives	10
<i>Aim</i>	10
<i>Objectives</i>	10
Chapter 2 Methods	11
Introduction	11
Oversight	11
Study participants and setting	11
Pre-index tests	11
<i>History and patient-completed questionnaires</i>	11
<i>Clinical examination and ultrasound</i>	12
Index tests	12
<i>Magnetic resonance imaging scan</i>	12
<i>Laparoscopy</i>	13
<i>Follow-up</i>	13
Assigning the cause of pain	14
Determining the role of magnetic resonance imaging by independent assessment	14
Reference standard	15
<i>Piloting the expert independent panel</i>	15
<i>Expert panel membership and meeting format</i>	16

Sample size	16
Data analysis	17
<i>Reliability estimates</i>	17
<i>Accuracy estimates</i>	17
Chapter 3 Diagnostic study results	19
Recruitment	19
Characteristics of participants	20
Reliability of the expert independent panel diagnoses	20
<i>Prevalence of conditions being the cause of pelvic pain or observed on magnetic resonance imaging scans or through laparoscopy</i>	22
<i>Comparison of observations on magnetic resonance imaging scans against a reference standard of observations made through laparoscopy</i>	23
<i>Comparison of observations made using magnetic resonance imaging against the expert panel reference standard for the cause of pelvic pain</i>	26
<i>The impact of magnetic resonance imaging on confidence and accuracy in making diagnoses</i>	26
<i>Accuracy of assessments of the cause of pelvic pain pre and post MRI and laparoscopy</i>	30
<i>Use and recommendations for therapeutic laparoscopy</i>	33
Chapter 4 Economic evaluation	37
Introduction	37
Methods	37
<i>Developing the model structure</i>	37
<i>Model assumptions</i>	43
<i>Data requirements</i>	44
Test accuracy	44
<i>Calculation of test-accuracy parameters</i>	46
<i>Further model parameters</i>	48
<i>Cost and resource data</i>	48
<i>Outcomes</i>	48
<i>Analysis</i>	49
Results	51
<i>Quality-adjusted life-year outcome</i>	51
Chapter 5 Discussion	61
Principal findings	61
Strengths and weaknesses of the diagnostic study	62
Strengths and weaknesses of the economic evaluation	64
Public and patient involvement	64
Interpretation of the findings	65
Implications for practice	65
Unanswered questions and further research	65
Acknowledgements	67
References	73
Appendix 1 Additional baseline data from the diagnostic study	79
Appendix 2 Additional results from the diagnostic study	89
Appendix 3 Economic evaluation sensitivity analyses	91

List of tables

TABLE 1 Summary of the evidence for accuracy of magnetic resonance imaging and laparoscopy in the investigation of CPP	4
TABLE 2 Radiological features identified on a pelvic MRI scan and the diagnostic criteria used	8
TABLE 3 Laparoscopic features identified via laparoscopy and criteria used	14
TABLE 4 Range of 95% CIs for varying prevalences	16
TABLE 5 Characteristics of participants	21
TABLE 6 Reliability of the EIP reference diagnoses	22
TABLE 7 Prevalence of conditions being the cause of pelvic pain or observed through MRI scans or laparoscopy	23
TABLE 8 Comparison of observations on MRI scans against a reference standard of observations made through laparoscopy	25
TABLE 9 Comparison of observations made using MRI against the expert panel reference standard for the cause of pelvic pain	27
TABLE 10 Impact of MRI on confidence in, and accuracy of, diagnoses made	31
TABLE 11 Difference between MRI and laparoscopy on confidence in, and accuracy of, diagnoses made	32
TABLE 12 Use and recommendations for therapeutic laparoscopy	34
TABLE 13 Independent gynaecologist determination of benefit from therapeutic laparoscopy	35
TABLE 14 Disease status of the women in the MEDAL study by treatment group	45
TABLE 15 Prevalence of structural causes stratified by treatment group	45
TABLE 16 Parameters used in the decision model	48
TABLE 17 Breakdown of the costs used in the economic analysis	49
TABLE 18 Quality-of-life values used to inform the QALY outcome	50
TABLE 19 Baseline deterministic cost-effectiveness results for MRI and laparoscopy at various cut-off points using the QALY outcome and a 6-month time horizon	51
TABLE 20 Baseline deterministic results for strategies on the cost-effectiveness frontier using the QALY outcome	52

TABLE 21 Baseline deterministic cost-effectiveness results for MRI and laparoscopy at various cut-off points, using the correctly diagnosed outcome and a 6-month time horizon	53
TABLE 22 Baseline deterministic results for strategies on the cost-effectiveness frontier, using the correctly diagnosed outcome	54
TABLE 23 Baseline deterministic cost-effectiveness results for MRI and laparoscopy at various cut-off points, using the correctly treated outcome and a 6-month time horizon	55
TABLE 24 Baseline deterministic results for strategies on the cost-effectiveness frontier, using the correctly treated strategy outcome	56
TABLE 25 Model results for the QALY outcome with variations in the time horizon of the analysis strategy	56
TABLE 26 Model results using variations in the prevalence of a structural cause of CPP	57
TABLE 27 Model results with a variation in the proportion of women with deep-infiltrating endometriosis who required a second therapeutic laparoscopy	58
TABLE 28 Model results with variation in the proportion of women with adenomyosis who were treated with a hysterectomy	58
TABLE 29 Recruitment by centre	79
TABLE 30 Demographic details of participants	80
TABLE 31 Participant characteristics	80
TABLE 32 Accuracy estimates for establishing the absence of any gynaecological cause against laparoscopy and the EIP assessment without MRI findings	89
TABLE 33 Accuracy estimates for establishing the absence of any gynaecological cause against laparoscopy and EIP with MRI findings	89
TABLE 34 Accuracy of independent gynaecologist's diagnosis before and after MRI	90
TABLE 35 Accuracy of treating gynaecologist's diagnosis before and after laparoscopy	90
TABLE 36 Distributions used in the PSA for all scenarios	91

List of figures

FIGURE 1 The MEDAL study flow chart	6
FIGURE 2 Patient pathway and data points through the MEDAL study	7
FIGURE 3 Flow chart of participants, showing the numbers who provided data at each stage	19
FIGURE 4 Relationships between the structural gynaecological causes	24
FIGURE 5 Accuracy of assessments of the cause of pelvic pain pre and post MRI and laparoscopy: ROC curves for comparison	28
FIGURE 6 Women with CPP who were eligible for laparoscopy, who received either MRI or laparoscopy	38
FIGURE 7 Model structure for the laparoscopy treatment arm for women who tested positive for a structural cause of CPP	39
FIGURE 8 Model structure for the MRI treatment arm for women who tested positive for a structural cause of CPP	40
FIGURE 9 Model structure for the laparoscopy treatment arm for women who tested negative for a structural cause of CPP	41
FIGURE 10 Model structure for the MRI treatment arm for women who tested negative for a structural cause of CPP	42
FIGURE 11 Receiver operating characteristic curves for laparoscopic and MRI diagnoses	46
FIGURE 12 Cost-effectiveness frontier for the baseline strategies, using the QALY outcome and the 6-month time horizon	52
FIGURE 13 Cost-effectiveness frontier for baseline strategies, using the correctly diagnosed outcome and a 6-month time horizon	53
FIGURE 14 Cost-effectiveness frontier for baseline strategies, using the correctly treated outcome with a 6-month time horizon	55
FIGURE 15 Cost-effectiveness acceptability curve showing the results of the PSA for a 6-month time horizon, for the strategies likely to be cost-effective across a range of willingness-to-pay thresholds for the QALY	59
FIGURE 16 Cost-effectiveness acceptability curve showing the uncertainty in the difference between costs and effectiveness (QALYs gained) for no testing and laparoscopy (cut-off 6) for 10,000 model runs using a 6-month time horizon	59

List of abbreviations

AUROC	area under the receiver operating characteristic	MEDAL	Magnetic resonance imaging for Establishing Diagnosis Against Laparoscopy
CI	confidence interval	MRI	magnetic resonance imaging
CPP	chronic pelvic pain	NICE	National Institute for Health and Care Excellence
EAU	European Association of Urology	PID	pelvic inflammatory disease
EIP	expert independent panel	PPSN	Pelvic Pain Support Network
EQ-5D-3L	EuroQol-5 Dimensions, three-level version	PSA	probabilistic sensitivity analysis
FS	fast spin	QALY	quality-adjusted life-year
GP	general practitioner	QoL	quality of life
ICECAP	Investigating Choice Experiments CAPability measure	RCOG	Royal College of Obstetricians and Gynaecologists
ICER	incremental cost-effectiveness ratio	ROC	receiver operating characteristic
IRRC	Independent Radiology Review Committee	SD	standard deviation
ISSP	International Society for the Study of Pain	SSC	Study Steering Committee
LR	likelihood ratio		

Plain English summary

Chronic pelvic pain (CPP) is pain in the pelvic and lower abdominal region that lasts for 6 months or longer. It can be caused by a gynaecological condition, such as endometriosis, or by bladder or bowel conditions. Most women will be offered laparoscopy, a keyhole surgery procedure to examine the inside of the abdomen, for treatment. Some other conditions can also be treated at the same time. In about half of all women, no obvious cause can be found for their pain.

Magnetic resonance imaging (MRI) scans can identify painful conditions, but we do not know whether or not they are better than laparoscopy. We performed a complex study in which nearly 300 women had a pelvic examination, an ultrasound scan, a MRI scan and then laparoscopy, and also completed symptom questionnaires. Gynaecologists considered all of this information and gave an opinion on the cause of pain for each woman. The study concluded that MRI scans could only correctly rule out a gynaecological condition in half of women judged to have no obvious cause, and missed half of the women who did have a gynaecological condition. Furthermore, MRI did not help to identify those women who could be treated during the laparoscopy; for every group of 1000 women with CPP having MRI scans, 369 women would then have an unnecessary laparoscopy because the MRI scans mistakenly showed a treatable cause, and 136 women would be incorrectly denied a laparoscopy, when actually there was a treatable cause that was not seen on the MRI scan. We do not think that all women should have a MRI scan using the currently available technologies.

Scientific summary

Background

Chronic pelvic pain (CPP) may be defined as pain in the pelvic and lower abdominal region, that lasts for ≥ 6 months. Idiopathic CPP is defined as CPP with an uncertain or unknown structural cause, after the exclusion of any other recognisable gynaecological pathology. Conditions such as endometriosis, fibroids, adenomyosis, cystic ovaries, adhesions and pelvic inflammatory disease are potential gynaecological causes, although CPP can have origins in the bowel, bladder, pelvic bones, muscles and fascia.

Symptoms experienced in CPP are variable and non-specific, so establishing a differential diagnosis can be hard. The Royal College of Obstetricians and Gynaecologists guidelines on the management of CPP [Royal College of Obstetricians and Gynaecologists. *Chronic Pelvic Pain, Initial Management (Green-top Guideline No. 41)*. London: Royal College of Obstetricians and Gynaecologists; 2012] provide a number of suggested initial investigations, including history, microbiological screening and vaginal examination, all based on low-quality evidence. If no cause of the pain is found, a diagnostic laparoscopy is often performed, although many conditions associated with CPP are not amenable to either laparoscopic diagnosis or treatment.

The aim of the Magnetic resonance imaging to Establish Diagnosis Against Laparoscopy (MEDAL) study was to assess if magnetic resonance imaging (MRI) can add value in the decision to perform a laparoscopy in women presenting with CPP. The important question was whether or not a normal MRI scan has a high enough negative predictive value to replace and avoid laparoscopy.

Objectives

1. To estimate the accuracy of post-MRI diagnoses with respect to (1) the absence of any observed condition or cause (i.e. idiopathic CPP); and (2) the main gynaecological conditions or causes of CPP as target conditions, using both post-laparoscopy diagnoses and expert independent panel (EIP) consensus (with and without incorporating MRI findings) as reference standards.
2. To determine the added value of MRI in a care pathway involving baseline history, clinical examination and ultrasound before performing laparoscopy.
3. To quantify the impact that pre-laparoscopy MRI scans can have on decision-making, with respect to triaging for therapeutic laparoscopy.
4. To perform a decision-analytic model-based economic evaluation determining the cost-effectiveness of MRI compared with laparoscopy.

Methods

We performed a comparative test-accuracy study with panel consensus for determining the reference standard.

Participants and setting

We recruited women aged ≥ 16 years, who had been referred to a gynaecologist, with CPP of at least 6 months' duration, and in whom there was an indication for diagnostic laparoscopy, from 26 UK hospitals. Women were excluded if they had had a hysterectomy, were pregnant, were unable to give written informed consent, definitely had a clinical indication for a MRI scan or if they had a previously established cause of CPP. Information concerning menstrual, obstetric and contraceptive history, pelvic examination and ultrasound scans was collected at baseline. Patient-completed questionnaires captured

domains such as (1) the location and severity of pain; (2) neuropathic, bowel-, urinary- and endometriosis-associated pain symptoms; (3) personality characteristics; (4) coping strategies; (5) depression; (6) prior sexual and physical abuse; and (7) general quality of life. Six months after diagnostic laparoscopy, data regarding treatments offered were collected and patient questionnaires were repeated.

Index tests

Diagnostic laparoscopy and MRI scans were the index tests. The MRI standard protocol comprised T1 and T2 axial, T2 sagittal, T2 coronal and T1-fast-spin (FS) axial, sagittal and coronal sequences. The observations were captured on a standardised report form and verified by a second radiologist who was blind to the initial MRI report. The diagnostic laparoscopy was performed under general anaesthetic in accordance with the gynaecologist's standard practice. When clinically indicated, additional therapeutic and diagnostic procedures were performed at the time of surgery.

To avoid verification bias, MRI was performed before diagnostic laparoscopy, and treating gynaecologists and investigators were kept blind to the MRI reports and images. To model the impact of using the MRI scan report, an independent gynaecologist reviewed each case. This gynaecologist was randomly selected from another hospital recruiting to the MEDAL study.

At several stages in the study path for each patient, the treating gynaecologist and the independent gynaecologist reviewed the diagnostic evidence available. To establish a possible diagnosis for each target condition (including idiopathic CPP), the gynaecologists indicated their level of certainty that the condition was causing the pelvic pain on a numerical rating scale, ranking from 0, meaning no chance, to 10, meaning certain.

Reference standard

The analysis considered two sets of reference standard. The first set identified was to assess the accuracy of MRI for observing a condition; the second was to assess the accuracy of MRI for identifying the condition(s) causing pelvic pain. The first set used findings on laparoscopy for the reference standard; the second used consensus from an EIP. The consensus process, the target condition definitions and the data presentation were standardised prior to the panel meetings, and reproducibility was assessed. The first-stage reference diagnosis was based on patient history and reported symptoms, clinical examination, ultrasound, laparoscopy and follow-up. For the second-stage reference diagnosis, the MRI scan report was provided. The panel members selected all conditions for which they were at least 50% certain that these were the underlying cause.

Sample size

A sample size of 250 women was chosen to address the primary research question of how many women could have avoided laparoscopy if MRI was routinely used in practice for diagnosing women with CPP. The study had over 90% power (at $p = 0.05$) to detect a reduction of 10% in the number of laparoscopies needed (i.e. from 100% down to 90%). With a prevalence of $\geq 30\%$ of idiopathic CPP, this sample size had the ability to reliably rule out a sensitivity or specificity of $< 70\%$ if the 'true' sensitivity or specificity was $> 80\%$.

Analysis

The prevalence of all gynaecological and non-gynaecological conditions in relation to whether or not they were believed to be causing pain (as determined by the reference standard established by consensus by the EIP) was calculated along with 95% confidence intervals (CIs) using binomial exact methods. The kappa value, alongside the 95% CIs, indicated the agreement between the three individual members of the EIP in their diagnoses.

The primary analysis involved calculations of sensitivity and specificity (and the associated 95% CIs) for the presence or absence of each structural gynaecological cause of pain; data were taken from the binary yes/no responses of the reference standard before and after the MRI data were revealed. Receiver operator

characteristic curves were also constructed for each test for each condition, using the certainty estimates provided by the treating gynaecologists and the independent gynaecologists, and the binary responses of the reference standard. The area under the receiver operating characteristic (AUROC) curve and 95% CI was estimated for each test. Differences between the AUROC estimates of each of the four diagnoses (pre and post laparoscopy by treating gynaecologist, and pre and post MRI by independent gynaecologist) were then calculated using a non-parametric approach for correlated data and are presented along with the 95% CIs. The same analysis was done to compare the responses after the MRI data were revealed to the EIP.

Cost-effectiveness evaluation

The cost-effectiveness for a range of scenarios for MRI and laparoscopy using different sensitivity and specificity values, as well as do-nothing (no MRI or laparoscopy) scenarios, was examined for women with undiagnosed CPP.

A decision-analytic model with a 6-month time horizon was developed in consultation with the clinical and methodological team. Two pathways for laparoscopy and MRI were constructed, with the cut-off value on the receiver operating characteristic (ROC) curve used to inform the sensitivity and specificity of each test being treated as a decision variable. Data collected in the MEDAL study to inform the prevalence, quality-adjusted life-year (QALY) and test-accuracy data were used to parameterise the economic model. Cost data were obtained from standard NHS sources.

A cost-utility analysis was undertaken based on the primary outcome of cost per QALY. The secondary outcomes of the cost per patient correctly diagnosed and the cost per patient appropriately treated were also considered as part of the wider cost-effectiveness analyses. The perspective of the analysis was the health-care provider perspective (the UK NHS) in a secondary care (hospital) setting. Given the 6-month time horizon, no discounting was applied. The results in this study are described using the incremental cost-effectiveness ratio (ICER). A range of one-way sensitivity analyses were undertaken to gain further insights into the impact of reasonable changes on key parameters in the model. A probabilistic sensitivity analysis was also conducted.

Results

Main findings of the test-accuracy study

The diagnostic study demonstrated that, compared with the reference standard diagnosis verified with laparoscopy, MRI had high specificity but poor sensitivity for observing deep-infiltrating endometriosis (3%), endometrioma (33%), adhesions (19%) and ovarian cysts (20%). Sensitivity was higher for detecting these conditions as a cause of pelvic pain, as categorised by the EIP, but not high enough to be useful. MRI correctly identified 56% (95% CI 48% to 64%) of women judged to have idiopathic CPP, but missed 46% (95% CI 37% to 55%) of women considered to have a gynaecological structural cause of CPP.

Magnetic resonance imaging added significant value in the care pathway for determining the cause of pain over and above the information gained by history, gynaecological examination and pelvic ultrasound (before performing diagnostic laparoscopy) in identifying deep-infiltrating endometriosis ($p = 0.006$) and endometrioma ($p = 0.02$), but not for other gynaecological structural causes or for identifying idiopathic CPP ($p = 0.08$).

Laparoscopy was significantly more accurate than MRI in diagnosing idiopathic CPP ($p < 0.0001$), superficial peritoneal endometriosis ($p < 0.0001$), deep-infiltrating endometriosis ($p < 0.0001$), and endometrioma of the ovary ($p = 0.02$) as the cause of pelvic pain. The accuracy of laparoscopy appeared able to rule in these diagnoses. Laparoscopy was not able to make accurate diagnoses that adhesions were the cause of pelvic pain, nor to diagnose adenomyosis.

Using MRI to identify women who require therapeutic laparoscopy would lead to 369 women in a cohort of 1000 women receiving laparoscopy unnecessarily, and 136 women who required laparoscopy not receiving it. A strategy of all women receiving laparoscopy would lead to 557 women who do not need laparoscopic treatment receiving it.

Main findings of the economic evaluation

Feeding these diagnostic performance data into a decision-analytic model to inform cost-effectiveness demonstrated that, for the majority of cut-off values used in the analysis, laparoscopy was found to be more effective, in terms of QALYs gained, than MRI. Laparoscopy was found to be increasingly cost-effective for longer time horizons. Extensive sensitivity analysis shows that, compared with laparoscopy, MRI is never a cost-effective option.

Conclusions

With respect to specific gynaecological conditions, MRI was accurate for endometrioma, adhesions, ovarian cysts and fibroids when compared with observations at laparoscopy. When compared with assignment of a cause for CPP symptoms by an expert panel, MRI was accurate for endometrioma, adhesions and fibroids. For determining the cause of pain, MRI added value in the care pathway over and above the information gained by history, gynaecological examination and pelvic ultrasound (before performing diagnostic laparoscopy) for the target conditions of deep-infiltrating endometriosis and endometrioma. The evaluation of non-gynaecological conditions suffered from a lack of agreement over verification of diagnosis by an expert panel, and this limits the useful interpretation of the data.

As a result of the extensive assumptions that were required to complete the economic analysis, the findings of the economic evaluation are insufficient on their own to inform practice. This study has shown that diagnostic laparoscopy is more cost-effective than MRI, although the importance of extending the time horizon and the cut-off point for the sensitivity and specificity of the tests has been demonstrated.

Strengths and limitations

Construction of an appropriate study designed to produce valid, reliable and generalisable results to guide decision-making proved challenging from the outset. The challenges included, among others, the variation in gynaecological practice, the multiplicity of target conditions with respect to the cause of CPP, the incongruence between observed conditions and attribution of the cause to CPP symptoms, the lack of consensus on radiology standards and the reporting of pelvic MRI for benign gynaecological conditions, the need to blind MRI results to avoid bias in the assessment of their value and the lack of objective reference standards for the various target conditions.

There is little evidence as to how diagnostic consensus panels should be operationalised. We piloted and then used a standardised approach to minimise variation in delineation of the reference standard, involving independent gynaecologists who provided the reference diagnosis for each case presented based on the consensus achieved through reviewing the structured summaries or full accounts of all the data collected. Understanding the difference between the presence of a condition and a target condition being the cause of pain proved to be the main challenge. During the course of the panel assessments, we undertook formal repeatability testing and found that the inter-rater agreement for assignment of gynaecological conditions as causes of CPP was good or very good, though it was poor or moderate for non-gynaecological conditions.

The strength of this model-based economic analysis is that it has utilised contemporary data from the MEDAL study and considered the cut-off point on the ROC curve to inform the optimum sensitivity and specificity for each test. However, it was limited to a 6-month time horizon for the data collection and does not capture the full long-term effects of successful testing and treatment for women with CPP. This means that the full patient benefit of laparoscopy and MRI has perhaps not been considered.

Implications for health care

Magnetic resonance imaging was dominated by laparoscopy in differential diagnosis of women presenting to gynaecology clinics with CPP. It did not add value to information, already gained from history, examination and ultrasound, about idiopathic CPP and various gynaecological conditions before considering a laparoscopy. To realise the little value that MRI did show for a few conditions, such as endometrioma and deep-infiltrating endometriosis, a value judgement about cost-effectiveness would have to be made, as the diagnostic gain from performing a MRI scan comes tagged with additional costs.

Recommendations for further research

The evaluation of tests for differential diagnosis in care pathways targeting various conditions for treatment requires methodological development (e.g. in areas of consensus science with respect to panel member numbers, sample size of cases for reliability). For CPP, the development of diagnostic prediction models incorporating biomarkers into baseline data may improve the use of diagnostic laparoscopy, but research will first be required to determine objective diagnostic criteria for the various target conditions and valid biomarkers for endometriosis.

Trial registration

This trial is registered as ISRCTN13028601.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Introduction

Background

Definition and prevalence of chronic pelvic pain

Chronic pelvic pain (CPP) may be defined as pain in the pelvic and lower abdominal region that lasts for 6 months or longer.¹ Idiopathic CPP is defined as CPP with an uncertain or unknown structural cause, that is, CPP that is unrelated to gynaecological organs, after the exclusion of other recognisable gynaecological pathology.² Symptoms include dysmenorrhoea (painful periods), dyspareunia (pain during sexual intercourse), dyschesia (painful bowel motions), dysuria (painful micturition) and constant or intermittent pelvic pain, which may or may not be related to the menstrual cycle.³

In primary care, the annual prevalence of CPP is 38 out of 1000 in women aged 15–73 years,⁴ a rate comparable with that of asthma (37/1000) and chronic back pain (41/1000).^{5,6} Unusually, no effective management policy exists for CPP. Only 20–25% of women respond to conservative treatment.⁷ CPP remains the single most common indication for referral to a gynaecology clinic, accounting for 20% of all outpatient appointments.⁸

Personal and societal costs of chronic pelvic pain

Chronic pelvic pain can have a substantial negative effect on a woman's quality of life (QoL). Women with CPP tend to report reduced general physical health scores than those without pain.^{9,10} Women with CPP describe loss, social isolation and effects on relationships. They have a high incidence of comorbidity, sleep disturbance and fatigue.³ They tend to cope outside the health system and usually do not see a health-care provider.⁸ Pain affects daily activities; around 18% of employed women in the UK take at least 1 day off work each year because of such pain.⁹ The annual direct cost of health care for women with CPP was estimated to be £158M, with a further £24M in indirect costs, in 1992.¹¹ More recent data for endometriosis estimated that the total annual societal burden of endometriosis-related symptoms was approximately £8.2B, with 1.5 million women affected.¹²

Pathogenesis of chronic pelvic pain

The pathogenesis of CPP is poorly understood. CPP may be a symptom of a single or multiple coexisting structural pathologies. The source of the pain can be either visceral (including the reproductive, urinary and gastrointestinal tracts) or somatic (including the pelvic bones, ligaments, muscles and fascia). The most recent taxonomy of pain from the International Society for the Study of Pain (ISSP) lists conditions such as endometriosis, fibroids, adenomyosis, cystic ovaries and pelvic inflammatory disease (PID) as pathological causes of CPP.¹³ Adhesions, frequently associated with endometriosis and pelvic congestion syndrome, are not defined within the ISSP taxonomy and, indeed, the relationship between CPP and pelvic vein insufficiency is controversial.¹⁴ The severity of pain may not be related to the severity of the underlying pathology, as illustrated by endometriosis, when the stage of disease is poorly correlated with reported pain.¹⁵ Pelvic pain may not be gynaecological in origin, however. Musculoskeletal dysfunction can include muscle laxity, manifesting as pelvic organ prolapse, muscle spasms (which can potentially lead to nerve entrapment) and trigger points (hyperirritable regions within skeletal muscle). Finally, pelvic pain may have origins in the bowel or the bladder. Irrespective of the underlying cause, women with CPP are more likely to have structural and functional changes to the central nervous system, and an increased risk of psychological distress and dysfunctional stress responses, than pain-free control women.¹⁶

A significant proportion of women with CPP appear to have no obvious underlying pathology identified during laparoscopy.¹⁷ Differences in the clinical thresholds for performing laparoscopy may vary over time or location, but data from the UK-based LUNA (Laparoscopic Uterosacral Nerve Ablation) trial¹⁷ showed that clinicians did not identify any condition at laparoscopy in 54% of participants. In these cases, the CPP

may have had an unknown cause that was not related to gynaecological organs and could be referred to as 'idiopathic'. The ISSP and European Association of Urology (EAU) define a subgroup of CPP, in which no obvious disease is found, to be CPP syndrome.^{13,18} Endometriosis-associated pain, vulval pain, bladder pain, primary dysmenorrhea and CPP with cyclical exacerbations are differentiated as subtypes of CPP syndromes by the ISSP. The EAU notes that, when pain is localised to a single organ, some specialists designate the pain accordingly, for example bladder pain syndrome, but when pain is localised to more than one site, the term 'CPP syndrome' should be used. Categorisation of the symptoms of idiopathic CPP syndrome into functional and psychological categories may then be appropriate. In a proportion of women, there may be a psychosomatic component, and psychological symptoms may be both causative and associative.

For this study, we used a working definition of idiopathic CPP with an uncertain or unknown structural cause after the exclusion of any other recognisable gynaecological pathology. This was derived by a nominal group method using an expert independent panel (EIP). A diagnosis of idiopathic CPP was arrived at once all other organic causes of pain were excluded by various diagnostic technologies or empirical treatment.²

Variation in the presentation of chronic pelvic pain

Symptoms experienced in CPP are variable and non-specific, so establishing a differential diagnosis can be hard. With this chronic condition, women present repeatedly over several years.¹⁹ It is possible in these cases that a previously diagnosed condition (e.g. endometriosis) has recurred or a new condition has developed (e.g. depression in a woman previously diagnosed with endometriosis). At present, there is wide variation in clinical practice concerning the diagnosis and management of CPP.²⁰ Women go from pillar to post, seeing several health professionals before eventually having their underlying condition identified. This wastes both the patient's time and NHS resources. The diagnosis of endometriosis may be delayed by over 8 years after first presentation with CPP symptoms,^{19,21,22} potentially demoralising the patient and missing the opportunity to improve their QoL and fertility through early effective treatment.

Diagnosis of the cause of chronic pelvic pain

A troublesome clinical issue is the lack of accurate tools to efficiently diagnose and direct cases of women.²³ The Royal College of Obstetricians and Gynaecologists (RCOG) guidelines on the management of CPP provide a number of suggested initial investigations, including history, microbiological screening and vaginal examination, all with weak evidence (levels B or C) for recommendations.²⁰ If no cause of pain is found, the first port of call is often to perform a diagnostic laparoscopy.

Clinical assessment and ultrasound

The order of tests in current practice can vary between clinicians and according to presentation, but the RCOG's clinical guidelines²⁰ suggest that a clinical history is taken and a vaginal examination is performed. When there is suspicion of PID or a potential need, screening for chlamydia and gonorrhoea in women aged < 25 years may be performed. A transvaginal ultrasound may be performed,²⁴ followed by a laparoscopy under general anaesthetic.

Laparoscopy

Diagnostic laparoscopy is a surgical procedure that allows a clinician to view the contents of the abdomen or pelvis to inform a diagnosis of CPP. A therapeutic laparoscopy is a surgical procedure used by a clinician to treat various conditions that may cause CPP. Endometriosis, pelvic adhesions, chronic PID and ovarian cysts are often observed via laparoscopy in women with CPP.⁸ In a cohort of 487 women recruited into a trial of neuroablation for CPP, 54% of women had no identifiable pathology at laparoscopy, whereas 31% had endometriosis, 5% had PID and 17% had adhesions.¹⁷ Approximately 11% had more than one finding. Those women with moderate to significant pathology were excluded from this trial. This is in broad agreement with other surveys.⁸

The role of laparoscopy in the differential diagnosis of chronic pelvic pain

There is evidence to suggest that deep-infiltrating endometriosis, bladder pain syndrome/interstitial cystitis²⁵ and irritable bowel syndrome^{26,27} are causally related to CPP, whereas, for adhesions, there is fair evidence. The associations between severe endometriosis and pelvic pain, and between endometriosis in general and infertility, are confirmed. However, there is little or no association between minimal endometriosis, pelvic adhesions or dilated pelvic veins and pain.²⁸ Debate remains over the relation between observed superficial peritoneal endometriosis and the extent of the laparoscopic finding,²⁹ or with the degree of reported pain.²⁸

A combination of the non-invasive tests for the diagnosis of endometriosis can sometimes provide insufficient or poor-quality evidence.³⁰ However, of all conditions associated with CPP, many are not amenable to either laparoscopic diagnosis or treatment. Of those conditions for which laparoscopy has a role, most are managed by gynaecologists, and this probably explains the common use of laparoscopy for evaluating women with CPP. Over 40% of laparoscopies are performed solely as a diagnostic test to identify the causes of CPP.³¹ The decision to perform a laparoscopy should be based on the history, physical examination and ultrasound scans suggesting a visceral origin.

The accuracy of laparoscopy in the diagnosis of chronic pelvic pain

The value of laparoscopy as a diagnostic tool for CPP has been considered in several papers,^{8,31,32} including a review of published reports of laparoscopically diagnosable conditions, in which an average of 61% of women undergoing laparoscopy for CPP had an identifiable pathology, compared with pathology in 28% of those without CPP.⁸ However, invisible (occult) endometriosis can be present in a seemingly normal peritoneum.³³ A 'negative' laparoscopy for some women may lead to them feeling disappointed that no diagnosis has been made³⁴ and disengage them from the care pathway.³⁵

However, a visual diagnosis of endometriosis during laparoscopy has been demonstrated to be unreliable. Only 54–67% of suspected endometriotic lesions are confirmed histologically, and 18% of women clinically suspected to have endometriosis have no evidence of endometriosis in biopsy samples.³⁶ A meta-analysis found that a positive finding through the use of laparoscopy will not be verified by histology in half of the cases (assuming a prevalence of 20%), although a negative laparoscopy is highly accurate for excluding endometriosis.³⁷ Other conditions identifiable through laparoscopy are listed in *Table 1*.

Magnetic resonance imaging in the diagnosis of chronic pelvic pain

Advances in imaging techniques suggest that magnetic resonance imaging (MRI) may be a useful non-invasive tool for diagnosing some conditions that cause CPP, such as adenomyosis⁴⁴ and deep-infiltrating endometriosis.⁴⁵ MRI has an established role in the diagnosis of adenomyosis and is currently the most accurate available non-invasive test in use.⁴⁶ The role of MRI in diagnosing small deposits of endometriosis²⁰ is less established. A Cochrane review reported that no imaging techniques met the criteria for a replacement or triage test for detecting superficial pelvic endometriosis.³⁰ Several factors contribute to this poor accuracy of MRI: non-pigmented lesions will not be hyperintense on T1 scans, small focal lesions may have variable signal intensity, plaque-like lesions are hard to delineate and adhesions cannot be directly identified, but are implied when the normal anatomy is distorted. Other conditions, such as fibroids and congenital uterine anomalies, may co-exist and are accurately diagnosed with MRI.⁴⁷

Magnetic resonance imaging is currently not recommended in guidelines nor used in routine practice for the investigation of CPP; the important question is whether or not a normal MRI scan has a high enough negative predictive value to replace and avoid laparoscopy. In CPP women, MRI may potentially have an important role in eliminating the need for surgery.

There are currently no agreed standardised protocols for MRI of the pelvis in the evaluation of pelvic pain. Radiologists determine the protocol used based on the clinical information and the suspected pathology, as well as personal preference and experience. Various bodies, such as the European Society of Urogenital Radiology (ESUR), have recommended protocols for pelvic assessment for various conditions,⁴⁸ but none specifically addresses CPP.

TABLE 1 Summary of the evidence for accuracy of magnetic resonance imaging and laparoscopy in the investigation of CPP

Target conditions	Diagnostic criteria	Diagnosis by	
		MRI	Laparoscopy
Endometriosis	Visualisation of endometriotic lesions with histological confirmation of biopsies from lesions ³⁷	Uncertain. MRI had a sensitivity of 85% and a specificity of 89% for detecting biopsy-proven endometriosis ³⁸ MRI had a sensitivity and a specificity for diagnosis of rectovaginal deep-infiltrating endometriosis of 55% and 99%, respectively. Sensitivity for other locations was considerably higher, for example, on uterosacral ligaments it was 85% ³⁹	Negative laparoscopy was accurate for excluding endometriosis (pooled LR -0.06, 95% CI 0.01 to 0.47) compared with biopsy, but positive findings were not as accurate (pooled LR 4.30, 95% CI 2.45 to 7.55) ³⁷
Adhesions	Visualisation, directly or by absence of movement between adjacent organs	Evidence from a single trial: sensitivity of 88% and specificity of 93% ⁴⁰	Gold standard
Chronic PID	Laparoscopic visualisation or histology from fimbrial mini biopsy	Ultrasonography represents preferable initial non-invasive diagnostic method. MRI sensitive for tubal, ovarian and pelvic abscesses ⁴¹	Can confirm tubo-ovarian abscesses and the presence of Fitz-Hugh–Curtis syndrome, but may fail to detect early disease or those with endosalpingitis only ^{42,43}
Adenomyosis	Presence of diffuse endometrial tissue in myometrium at post-hysterectomy biopsy	MRI had a pooled sensitivity of 77% and a specificity of 89% against histological diagnosis ⁴⁴	Uncertain. May observe bulky uterus

CI, confidence interval; LR, likelihood ratio; MRI, magnetic resonance imaging.

Treatment of chronic pelvic pain

As diagnoses emerge through careful history and examination and directed investigation, so may treatment strategies. These should be tailored to the needs of individual women, whatever the cause. A multidisciplinary approach is considered ideal for achieving this. Chronic symptoms need long-term management and a multimodal approach.

Some clinicians prefer that a diagnostic and therapeutic investigation is performed as part of a single procedure – sometimes referred to as ‘see-and-treat’ laparoscopy.^{49,50} For example, adhesiolysis may be performed in what started as a diagnostic procedure.⁵¹ Depending on the severity and corresponding patient consent, endometriosis may be treated at the time of diagnosis by either electrocoagulation, laser vaporisation or excision.⁵² Endometrioma and benign ovarian cysts are usually excised.⁵³ The type of treatment offered may depend on the extent of the disease and surgical expertise. Surgical removal of deep-infiltrating endometriosis is highly specialised and, in the UK, undertaken in accredited specialist centres. Preoperative MRI scans can aid the planning of this complex surgery.

Other conditions observed at laparoscopy are not amenable to immediate laparoscopic treatment, for example myomectomy for subserosal fibroids.⁵⁴ Few surgeons attempt uterine-sparing surgery for adenomyosis, as a result of it frequently being diffused throughout the myometrium.

A diagnostic laparoscopy may also provide the opportunity to undertake other investigations, such as a cystoscopy, if symptoms are suggestive of bladder or intrauterine conditions. Some gynaecologists may also use the cover of a general anaesthetic to perform investigations capable of being performed as an outpatient procedure, such as hysteroscopy or insertion of a levonorgestrel-releasing intrauterine system.

Medical treatments are also options for CPP, irrespective of surgical intervention. The combined oral contraceptive pill is used post laparoscopy for prevention of the recurrence of endometriosis, as are long-acting reversible contraceptives.^{55–57} Gabapentin is used for women with no identifiable cause of pain, although evidence for its efficacy is limited.⁵⁸

Overview of the Magnetic resonance imaging for Establishing Diagnosis Against Laparoscopy study

The MRI for Establishing Diagnosis Against Laparoscopy (MEDAL) study recruited women from 26 UK hospitals (as listed in *Appendix 1, Table 29*). An outline of the test-accuracy study design is shown in *Figure 1*. MRI was undertaken before laparoscopy, but the resulting report and images were not to be provided to the gynaecologist, unless there was a critical finding, such as suspected malignancy. This blinding was necessary in order not to distort clinical practice and avoid verification bias arising from knowledge of this index test. A diagnostic laparoscopy was performed and, together with information from the history, examination and ultrasound, produced a post-laparoscopy diagnosis (*Figure 2*). Information was collected from those who were not eligible for the test-accuracy study. Follow-up information was elicited directly from participants at 6 months.

An economic evaluation was performed to establish the cost-effectiveness of MRI as an alternative to laparoscopy for the sensitivity and specificity values of the various target conditions.

Rationale for the Magnetic resonance imaging for Establishing Diagnosis Against Laparoscopy study

There is a perception that diagnostic laparoscopy, an invasive, expensive and potentially risky procedure, is used far too frequently in the NHS, as a significant proportion of women have no pathology identified.⁵⁹ The procedure is associated with an approximate 3% risk of minor complications (e.g. nausea and vomiting, shoulder-tip pain) and a 0.24% risk of unanticipated injury causing major complications (e.g. bowel perforation), resulting in two-thirds of women requiring laparotomy.^{59–61} There is an estimated risk of death of 3.3 to 8 per 100,000 women,^{59,61} and payments in medical negligence cases totalled £24.3M in one survey published in 2000.⁶²

Magnetic resonance imaging is easily accessible, less invasive and cheaper than surgery. In some circumstances, MRI may add a diagnostic benefit over laparoscopy, such as in the cases of severe deep-infiltrating endometriosis and adenomyosis. MRI may also be used to triage cases, so that therapeutic laparoscopy could be better planned. The aim of the MEDAL study was to determine the proportion of women for whom MRI could remove the need for a laparoscopy or, in other words, for whom MRI could be a replacement for laparoscopy.

For MRI to be used in routine clinical practice for the evaluation of CPP, imaging protocols and reporting need to be standardised to allow homogeneous MRI performance between centres and to limit acquisition and interobserver variability. Currently, there are no routinely used standardised reporting guidelines for CPP in the UK.⁶³ Guided by the MEDAL study, members of the study panel have proposed a template for the standardised reporting of CPP.⁶³

Following various treatment options, such as the oral contraceptive pill or hormone treatment, some women will undergo laparoscopy resulting in 50% of women having a negative laparoscopy with no pathological cause of pain identified.⁸ Although a negative laparoscopy may fail to identify the pathology, it may have a reassuring effect on the patient. A laparoscopy may also have the potential to be a 'see-and-treat' laparoscopy if superficial peritoneal endometriosis is observed; MRI would add no benefit in that case.

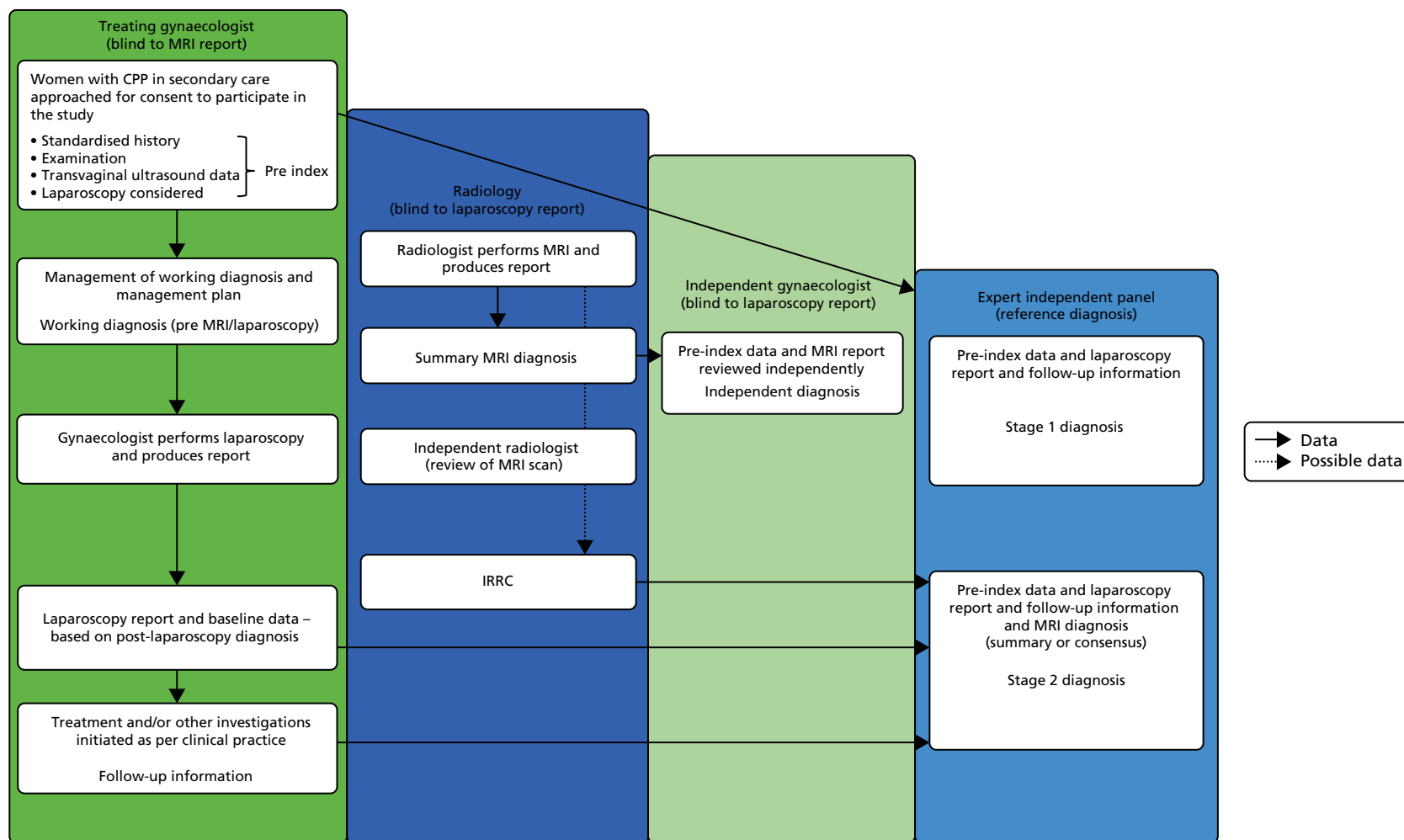


FIGURE 1 The MEDAL study flow chart. IRRC, Independent Radiology Review Committee.

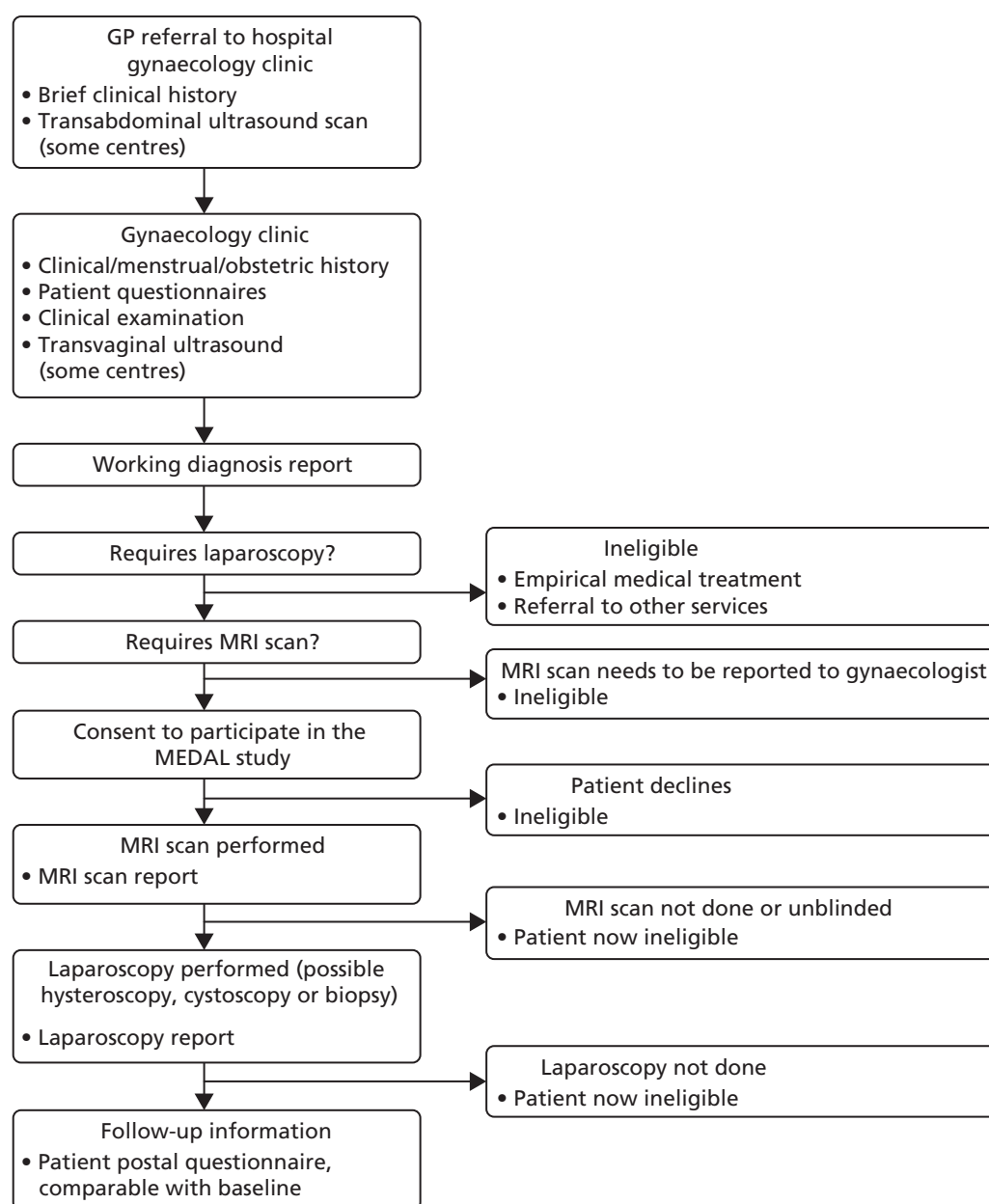


FIGURE 2 Patient pathway and data points through the MEDAL study. GP, general practitioner.

The choice of study design

Test-accuracy studies are designed to generate measures of accuracy by comparison of the index test with a reference standard, a test that confirms or refutes the presence or absence of disease beyond reasonable doubt. Classical test-accuracy studies require that the target condition is independently verified by the reference test, which must provide a definitive diagnosis, be applicable in all cases and preferably be performed alongside the index test. Complete verification of the presence or absence of a target condition is essential to reduce bias and to maximise the statistical power of the study. By comparing the index test with the reference standard, the result of the index test can be categorised as a true positive, a false positive, a true negative or a false negative. Measures of index test accuracy, including sensitivity, specificity, positive and negative likelihood ratios, and diagnostic odds ratios, can be computed.⁶⁴

Verification of the reference standard is frequently dichotomised as disease present/disease absent. In the MEDAL study, we aimed to determine the accuracy of MRI in the differential diagnosis of the causes of CPP.

There is a range of potential target conditions and, correspondingly, of reference diagnoses. MRI will visualise the various conditions with different degrees of accuracy (*Table 2*). To determine the sensitivity of MRI for each pathology would require a large number of participants, to accommodate the low prevalence of some conditions. Furthermore, some pathologies are not independent of each other and could frequently be concurrently observed; for example, endometriosis can give rise to adhesions from fibrotic tissue. Therefore, the principal research question is how frequently does a MRI scan correctly predict the reference diagnosis, thereby removing the need for a laparoscopy or, conversely, what diagnoses cannot be accurately ruled out by a MRI scan and require laparoscopic investigation?

In the context of the MEDAL study, there were a number of target conditions to be considered, not all of which had a perfect reference standard of identification against which to make a differential diagnosis. There was a risk of partial or differential verification of the underlying causes, with their inherent biases.

TABLE 2 Radiological features identified on a pelvic MRI scan and the diagnostic criteria used

Radiological feature	Criteria used in analysis to indicate the presence of a target condition	Reference or justification
Anatomy of uterus, ovaries, adnexal structure, bladder and bowel	Not used for analysis	
Uterine size, appearance, JZ thickness, adjacent myometrial thickness, endometrial thickness	Adenomyosis is diagnosed in two ways: 1. If JZ is > 50% of adjacent myometrium thickness (including JZ) 2. Thickness of JZ is > 13 mm, in either the anterior or the posterior wall or fundus	
Presence of fibroids, location, number and maximum size	Presence of fibroids reported, irrespective of location, number or size	Fibroids are seen on the MRI scan
Ovarian size, ovarian cysts, and, if present, their size and signal intensity	Endometrioma of the ovary is diagnosed if the ovary had a cystic area or deposit with: <ul style="list-style-type: none">high or intermediate signal intensity on T1, T2 and T1-FS imagesonly a high signal intensity on T1, T2 and T1-FS images[Cysts (other than follicles) = present, T2 = high and T1, T1-FS = low or intermediate]	
Presence and location of other masses	<ul style="list-style-type: none">Diagnosis of deep-infiltrating endometriosis if masses or thickening are seen in the bowel or bladder; andmasses have a high or intermediate signal intensity on T1 and T1-FS imagesmasses have a high signal intensity on T1 and T1-FS images	Bazot <i>et al.</i> , ³⁹ Bazot <i>et al.</i> ⁶⁵ and Chamié <i>et al.</i> ⁶⁶
Presence of free or loculated fluid	Diagnosis of PID if either free fluid or tubal masses are seen, with low or intermediate signal intensity on T1, T2 and T1-FS images	
Adhesions	Diagnosis of adhesions if reported as: <ul style="list-style-type: none">clearly seenclearly seen or suspected	
Bladder status, including wall thickness if abnormal	Not used for analysis	
Presence of small bowel in pelvis, location and description	Not used for analysis	
Other observations; for example, lumbar spine abnormalities	Not used for analysis	

FS, fast spin; JZ, junctional zone.

There are several proposed study designs that overcome the problems of an imperfect reference diagnosis.⁶⁷ In the absence of a single 'gold standard' test, the results of several imperfect tests or observations can be combined to create a composite reference standard. These can be combined according to a predefined algorithm or a consensus diagnosis obtained from considering all the information.

We employed a panel (consensus) diagnosis in which a group of experts determined the presence or absence of the target condition based on several sources of information captured along the patient pathway. This was an acceptable way of addressing the problem of achieving a diagnosis from multiple sources of information and of subjective assessment of that information, as a diagnosis is achieved by consensus.⁶⁸ A final diagnosis can be obtained for all women and the panel method reflects the clinical reality, in which several items of information are synthesised by the clinician. The results of this methodology can be viewed as being generalisable to clinical practice.

To establish whether or not MRI can replace laparoscopy, a paired design was employed, in which both tests were compared with the reference standard. As sensitivity and specificity can vary across subgroups, the two tests and the reference standard are best performed in the same population.⁶⁸ It is feasible to perform MRI and laparoscopy in women with CPP and, although MRI does not interfere with laparoscopy, the paired design is preferable to a randomised trial.⁶⁹ As MRI would precede laparoscopy, it can also be viewed as a triage test, directing only women with a specific condition towards a laparoscopic confirmation or operative laparoscopic procedure. A fully paired study design is also appropriate for this type of diagnostic situation.

The certainty with which gynaecologists have made their diagnoses will be measured and compared as a secondary measure of diagnostic efficacy. In order to be clinically effective, tests should contribute to the diagnostician's decision-making,⁷⁰ for example, by changing a differential diagnosis, strengthening an existing hypothesis or simply reassuring the clinician. Although accuracy is the chief concern, the extent to which clinicians make use of test results also relies on their confidence that the test has contributed usefully to a diagnosis. The MEDAL study therefore evaluated the diagnostic impact of MRI and of laparoscopy by conducting a:

1. before-and-after comparison of diagnostic certainty for having diagnosed the cause of CPP
2. before-and-after comparison of diagnostic certainty for the leading diagnosis
3. before-and-after comparison of the number of differential diagnoses considered per patient
4. retrospective survey of the test's perceived usefulness.

On the basis of clinical history, examination and ultrasound findings only, treating gynaecologists were asked to state whether or not a pathological cause for CPP was identified, and to express their certainty regarding this decision. They were also asked to list their differential diagnoses, to state whether each was thought to be a cause of CPP and to express their certainty regarding these opinions. After the laparoscopy was performed, clinicians were asked for a revised differential diagnosis and associated certainty using identical questions. A comparison between pre- and post-laparoscopy diagnostic certainty was made, primarily to determine the utility of laparoscopy for identifying a pathological cause of CPP, and secondarily to compare changes in the certainty surrounding the leading differential diagnosis. The number of differential diagnoses considered before laparoscopy was also compared directly with the number considered after the test results were known.

An identical process was followed with independent non-treating gynaecologists (who were blind to the laparoscopic findings) who used MRI to arrive at a diagnosis. Pre-MRI diagnoses and associated certainty based on the same clinical history, examination and ultrasound findings were compared with post-MRI revised diagnoses and certainties. Independent gynaecologists were also asked whether or not they believed that the patient required a laparoscopy.

The gynaecologist's subjective assessment of the usefulness of MRI or laparoscopy was evaluated after disclosure of the relevant test results, by asking clinicians to select one of four statements that best reflected their perception of the contribution of the test to each case. These statements were initially

drafted with reference to those used in published before-and-after diagnostic confidence studies,^{71–73} and modified for their relevance to a CPP diagnosis following consultation with eight practising gynaecologists.

Aim and objectives

Aim

The aim was to assess if MRI could add value to the decision to perform a laparoscopy in women presenting with CPP. In doing so, we wanted to determine the number of women for whom MRI is sufficiently accurate to avoid the need for a laparoscopy in the investigation of CPP, following evaluation of the presenting characteristics with a detailed history, clinical examination and ultrasound.

Objectives

- To estimate the accuracy of post-MRI diagnoses with respect to:
 - the absence of any observed condition or cause, that is, idiopathic CPP
 - the main gynaecological conditions or causes of CPP as target conditions, using post-laparoscopy diagnoses and EIP consensus (with and without incorporating MRI findings) as reference standards.
- To determine the added value of MRI in a care pathway involving baseline history, clinical examination and ultrasound before performing laparoscopy.
- To quantify the impact that a pre-laparoscopy MRI scan can have on decision-making with respect to triaging for therapeutic laparoscopy.
- To determine the cost-effectiveness of MRI compared with laparoscopy.

Chapter 2 Methods

Introduction

We conducted a test-accuracy study in a prospective cohort of women with CPP symptoms, subjected them to thorough investigations, including our index tests (MRI scan and laparoscopy), and sought expert panel consensus to establish a reference diagnosis. We performed a comparative test-accuracy study with panel consensus for determining the reference standard.

Oversight

The study was conducted in accordance with the protocol, which received a favourable ethics opinion from East Midlands (Nottingham 1) Research Ethics Committee (reference number 11/EM/0281). NHS trust research governance approval was obtained from 26 recruiting hospitals in the UK. The study sponsor was the Queen Mary University of London. Independent oversight for the MEDAL study was requested by the funding body; however, how this oversight should be arranged or what it should entail were unclear, as the study was not considered to be distinct from that of randomised controlled trials. For the MEDAL study, the Data Monitoring Committee was defined as a subgroup of the Study Steering Committee (SSC). All members were independent of the MEDAL study management group. The chairperson of the SSC (along with the lay member) was not included in the Data Monitoring Committee.

Study oversight was provided by an independent steering committee and an independent data monitoring committee (see *Acknowledgements* for membership).

Study participants and setting

Any woman aged ≥ 16 years, who had been referred to a gynaecologist with CPP of at least 6 months' duration, was potentially eligible. Women were assessed by a gynaecologist, which involved a standardised assessment and a clinical examination. If clinically indicated, women were referred for an ultrasound scan. If a laparoscopy was indicated and the patient wished to proceed with the laparoscopy, and was prepared to have a MRI scan, eligibility to participate in the study was confirmed. Women were excluded if they had had a hysterectomy, were pregnant or were unable to give written informed consent, or if they definitely had a clinical indication for a MRI scan, based on examination, history or ultrasound. Women were also excluded if they had an identifiable cause of CPP for which treatment could be initiated without a laparoscopy. Eligible women who provided written informed consent were committed to the study by telephone registration service at Birmingham Clinical Trials Unit (University of Birmingham, Birmingham, UK).

Pre-index tests

History and patient-completed questionnaires

A structured assessment template for use by all gynaecologists was provided, which first collected data from the participant and then allowed the clinician to create a clinical record. The assessment incorporated (1) demographic details; (2) menstrual, contraceptive and obstetric history; (3) tobacco and alcohol use; and (4) previous investigations and treatments for pelvic pain. Visual analogue scales were used to elicit scores (from 0 to 10) for pain symptoms of (1) dysmenorrhea (including pain occurring pre menses, during menstruation, post menses and mid-cycle); (2) dyspareunia (at the point of penetration, deep pain, burning pain, post-coital pain); (3) dysuria (full bladder, micturition); (4) muscle/joint pain in the pelvis; and

(5) backache and migraine. A body map to illustrate the location of pain was also provided. The assessment also included standardised patient-completed questionnaires, previously validated in either women with pelvic pain or other clinical symptom groups. The domains captured by the questionnaire components of the assessment were:

- neuropathic pain symptoms, measured using the short-form McGill Pain Questionnaire⁷⁴
- bowel symptoms, measured using the Rome III diagnostic criteria⁷⁵
- urinary problems, measured using the Pelvic Pain and Urgency/Frequency (PUF) Patient Symptom Scale⁷⁶
- sexual activity, measured using the Sexual Activity Questionnaire⁷⁷
- personality characteristic, measured using the Big Five Inventory scale⁷⁸
- psychological coping strategies, measured using the Pain Catastrophizing Scale⁷⁹
- endometriosis-related pain, measured using the Short Form Endometriosis Health Profile Questionnaire (EHP-5)⁸⁰
- risk of depression, measured using the Patient Health Questionnaire depression module (PHQ-2)⁸¹
- sexual and physical abuse history, measured using the Sexual and Physical Abuse History questionnaire⁸²
- QoL and well-being, measured using the EuroQol-5 Dimensions, three-level version (EQ-5D-3L) instrument and the Investigating Choice Experiments CAPability (ICECAP)-A measure.^{83,84}

Clinical examination and ultrasound

Each woman underwent a clinical examination by the recruiting gynaecologist. A data collection form helped to direct the examination, which aimed to describe the location of any abdominal tenderness, and any tenderness, erythema, discharge or lesions on the external genitalia. A bimanual examination aimed to identify vaginismus, bladder wall or uterosacral ligament tenderness, cervical excitation, uterine size, position, contour, consistency and mobility, ovarian tenderness, mobility and presence of masses. A speculum examination of the vagina was used to identify prolapse, vaginal atrophy, polyps and cervicitis, and also allowed swabs to be taken if indicated and if the woman consented. A urine sample was provided to perform a dipstick test to detect urinary tract infection.

The intention was for all women to have a detailed ultrasound scan, in accordance with the study protocol, scanning by both the transabdominal route on a full bladder and then the transvaginal approach on an empty bladder. However, in some centres, the ultrasound was scheduled with a sonographer prior to the first appointment with the gynaecologist. In these cases, it was not possible to repeat the ultrasound, and an incomplete data set for the study was extracted from the report. The intention was to describe the uterine size by both approaches, and also to examine the uterine position, myometrial appearance and endometrial width, as well as ovarian size, location and presence of ovarian cysts and, finally, bladder wall thickness, all by transvaginal ultrasound. Further information sought included the presence of free or loculated fluid and organ- or site-specific tenderness or restricted mobility.

Index tests

Magnetic resonance imaging scan

Women were asked to fast for 2–4 hours prior to the scan and not to empty their bladder immediately prior to the scan. Antiperistaltic agents were contraindicated, but antispasmodic drugs were acceptable.

The MRI standard protocol comprised T1 and T2 axial, T2 sagittal, T2 coronal and T1-fast-spin (FS) axial, sagittal and coronal sequences, using standardised anatomical landmarks, slice thicknesses and field of view. If unexpected abnormalities were detected, additional sequences, potentially using the contrast media gadolinium, could be added to the scanning protocol to enable further clarification and assessment.

To ensure blinding of assessments, a MRI scan was performed before diagnostic laparoscopy, and treating gynaecologists and investigators were kept blind to the MRI reports and images. To identify radiological

parameters for inclusion in the MRI report, a two-generational Delphi survey with an expert panel of 28 radiologists specialising in gynaecological MRI from across the UK was undertaken.⁶³ The MRI report collected data described in *Table 2*.

The radiologists' assessments of the quality of the scan were captured, and they were asked to summarise the observations and provide a radiological diagnosis in free-text form.

The MRI scan was also independently reported by a second radiologist who was blinded to the initial MRI report. This would be a radiologist involved in reviewing scans for the study, but based at a different site from the initial reporting radiologist. Where the local and independent radiologists agreed on the summary diagnosis, the findings were accepted as a consensus and provided the 'post-MRI diagnosis' data for analysis. Lack of agreement between local and independent radiologists prompted a review by the Independent Radiology Review Committee (IRRC), made up of three experienced radiologists who were not involved in imaging participants in the study. The IRRC agreed a consensus MRI report. The finalised report was provided to the independent gynaecologist for determining the post-MRI diagnosis, and to the EIP for the second stage in the reference diagnosis.

The diagnostic criteria used for the analysis of the accuracy of MRI are given in *Table 2*. These were defined by the lead radiological investigator and confirmed by a second radiologist otherwise not involved in the study. There were potential variations in diagnosis, for example in relation to whether an intermediate signal from a T1 and T1-FS image of a bowel mass was indicative of deep-infiltrating endometriosis; therefore, the different diagnostic criteria were considered in alternative or sensitivity analyses.

The MRI reports were not provided to the gynaecologist unless unexpected significant findings were identified by the local reporting radiologist. We prespecified the circumstances in which this may happen, such as identification of an unexpected cancer, abscess or non-gynaecological abnormality requiring immediate attention. In the case of unblinding, the participant was excluded from the diagnostic accuracy study and managed appropriately.

Laparoscopy

The diagnostic laparoscopy was performed under general anaesthetic in accordance with the gynaecologist's standard practice. After a pneumoperitonium was established, a laparoscope was introduced to visualise the abdominal and pelvic structures, and any pathology. Laparoscopy was performed by an experienced gynaecologist who was capable of identifying all potential target conditions. Surgical findings were reported using a standardised proforma, which collected the information described in *Table 3*.

When clinically indicated, additional procedures were performed at the time of surgery. Information was collected on procedures, including ablation or excision of superficial endometriosis, ovarian cystectomy, adhesiolysis, hysteroscopy (noting the presence of fibroids or the insertion of the levonorgestrel-releasing intrauterine system), endometrial biopsies and histological confirmation of pathology and cystoscopy. If bladder biopsies were collected, the histological confirmation of pathology was recorded.

Follow-up

Some of the target conditions may respond well to treatment; therefore, information was collected at 6 months following laparoscopy, using a postal questionnaire. The questionnaire contained all of the elements of the initial patient questionnaire, including the visual analogue scales, validated questionnaires and questions regarding further tests and treatments, but omitted the demographic and history questions, as the answers to these questions would not have changed. Women were sent reminders if they failed to respond to the initial questionnaire. The information gained from follow-up was considered by the EIP when assigning the reference standard.

TABLE 3 Laparoscopic features identified via laparoscopy and criteria used

Laparoscopic feature	Criteria used in analysis to define the observation of the target condition
Presence, location and severity of endometriosis	Presence of endometriosis in any location, categorised as either superficial or deep. Biopsies may have been taken for histological confirmation, but the results were not considered for the criteria. AFS grading was used ⁸⁵
Presence, location and appearance (filmy, dense/vascular or an absent plane) of adhesions	Presence of adhesions in any location, irrespective of appearance
Presence of adenomyosis	Presence of adenomyosis observed. It is acknowledged that adenomyosis cannot be directly observed, but other features might be suggestive
Presence, location, type (simple, dermoid, cancerous, other) and size of ovarian cysts	Presence of any ovarian cysts, regardless of type
Presence, location and maximum size of fibroids	Presence of any fibroids, regardless of location. It is acknowledged that submucosal and small intramural fibroids cannot be directly observed, but other features may be suggestive or they may have been observed by hysteroscopy
Presence of PID	Presence of features indicative of PID
Pelvic congestion syndrome	Observation of dilated pelvic veins
AFS, American Fertility Society.	

Assigning the cause of pain

At several stages in the study path for each patient, the treating gynaecologist and the independent gynaecologist reviewed the diagnostic evidence available. To establish a possible diagnosis for each target condition (including idiopathic CPP), they indicated their level of certainty that the condition was causing the pelvic pain on a numerical rating scale, ranking from 10 being certain (expressed as a 99 in 100 chance), through 5 being a fairly good possibility (a 5 in 10 chance) to 0, considered to be no chance or less than a 1 in 100 chance.

The first condition was the absence of structural cause, defined as idiopathic or unknown. There then followed seven structural gynaecological causes (superficial peritoneal endometriosis, deep-infiltrating endometriosis, endometrioma of the ovary, adhesions, ovarian cysts, adenomyosis, fibroids), together with the option to specify another gynaecological cause. The non-gynaecological causes offered were psychological or psychosexual, gastrointestinal, urinary, musculoskeletal, neurological or another pathological cause, with the requirement to specify the diagnosis further. Definitions for idiopathic CPP, superficial peritoneal and deep-infiltrating endometriosis, adhesions and pelvic congestion syndrome were provided.

Determining the role of magnetic resonance imaging by independent assessment

The treating gynaecologists were kept blind to the MRI scan report so that they could make an unbiased assessment of the laparoscopic findings. If MRI was found to be useful in the MEDAL study, an MRI scan would precede laparoscopy in the investigative pathway in the future. In order to model the impact of using the MRI scan report, an independent gynaecologist reviewed each case. This gynaecologist was randomly selected from another hospital recruiting to the MEDAL study, so was familiar with the study aims, processes and data collection. The independent gynaecologist was provided with all the data collection forms for the pre-index tests and clinical history, together with the laparoscopy report. The information was anonymised. The independent gynaecologists were asked to complete the certainty scales for each condition, which

constituted the pre-MRI independent diagnosis and should be highly correlated with the recruiting gynaecologists' diagnoses post laparoscopy, being derived from the same information. The independent gynaecologists were then asked to read the MRI scan report and repeat the process, completing the certainty scales. There were three extra hypothetical questions at this second post-MRI stage: if the gynaecologist was treating that particular patient, would they have scheduled a laparoscopy (having reviewed the MRI scan report), if they would have scheduled a therapeutic laparoscopy and, if so, would they anticipate that the diagnostic and therapeutic elements would be performed as part of a single procedure?

Reference standard

The analysis considers two sets of reference standard. The first identified was to assess the accuracy of the MRI scan for observing a condition, and the second was to assess the accuracy of the MRI scan for identifying the condition(s) causing pelvic pain. The first set used findings from laparoscopy for the reference standard, and the second used consensus from an EIP. It was expected that not all women who were observed to have each condition would have it categorised as being the source of pelvic pain.

Piloting the expert independent panel

There is little evidence on how panels should be convened or how information should be presented. Given the amount of information to be considered by the panel and the associated review time, the study management group, in conjunction with the EIP chairperson, considered the use of a summary report. To ensure that the diagnosis based on the summary report was the same as the diagnosis based on all of the data collection forms (initial patient assessment, including patient questionnaire, clinical examination, ultrasound, post-laparoscopy report and, finally, the MRI scan report), a pilot EIP panel was convened to pilot and review the process. The pilot EIP panel consisted of three clinicians not involved in the study recruitment.

Ten cases were presented to the three members. For the first five cases, two of the three members used the summary report and the third member used the data forms. For the last five cases, two members, who had previously used the summary report, used the data collection forms. For each case, the members defined a single or multiple diagnosis. Results from the meeting indicated that the summary report and data collection forms allowed the members to produce the same diagnosis. Data collection forms were preferred, or used in conjunction with the summary report, when cases were considered to be complex (e.g. when a patient had multiple conditions and various confounding factors).

To elicit a diagnosis, a form with a list of the target conditions was provided. This was comparable to the form used by the treating and independent gynaecologist, except that the certainty scale was not used and simple yes/no answers were allowed. During the pilot EIP meeting, members were allowed to record multiple causes. This was considered unsatisfactory as a reference standard, as they would complicate accuracy assessments, so the initial guidance regarding the first eight cases presented to the main EIP members was to diagnose what they believed to be the single main cause of pain. The process was performed in two stages: once with all of the data, except those from the MRI report, and then a second time, including the data from the MRI report. The first-stage diagnosis avoided incorporation bias from knowledge of the MRI scan results, whereas the second-stage diagnosis provided information to determine the added value of the MRI scan. Two further yes/no questions were posed to the panel: (1) does the laparoscopy add anything to your diagnosis that was not available from other investigations and (2) could this laparoscopy have a therapeutic purpose?

Feedback from these eight cases suggested that, in some instances, it was necessary to select more than one condition, given that two or more conditions could equally be the cause of pain. The criteria and guidance were changed so that EIP members could select any clinically appropriate conditions they believed to be the cause of pain, provided that they were at least 50% certain. The diagnosis form still allowed only yes/no answers, but multiple reasons could be given. The first eight cases were reviewed again at a later meeting using the revised criteria.

Expert panel membership and meeting format

The panel membership consisted of 15 consultant gynaecologists who were not involved in the study recruitment. Each face-to-face meeting was made up of three members who provided the reference diagnosis for each case presented based on the summary report of the data, with the individual data forms available if required.

The first-stage reference diagnosis was based on patient history and reported symptoms, clinical examination, ultrasound, laparoscopy and follow-up. For the second-stage reference diagnosis, the MRI scan report was provided. For both stages, each EIP member individually recorded the conditions that they were more than 50% certain were the cause of pain, and addressed the two questions regarding the potential gain from the laparoscopy, prior to a group discussion. The meeting chairperson documented the discussion and how agreement was reached, and the final consensus diagnosis constituted the reference standard.

Sample size

A sample size of 250 women was chosen a priori to address the primary research question of determining the proportion of women for whom MRI is sufficiently accurate to replace laparoscopy in the investigation of CPP, following evaluation of the presenting characteristics. With this sample size, the study was anticipated to have > 90% power (at $p = 0.05$) to detect a reduction of 10% in the number of laparoscopies needed (i.e. from 100% down to 90%). This difference would be cost-effective if laparoscopy was at least 10 times more expensive than MRI. Estimates used at the start of the study placed laparoscopy at being 7.4 times more expensive than MRI (£1274 vs. £173); however, these NHS tariffs may not necessarily reflect the true cost of the total investigative pathway, which will be estimated through primary data collection as part of this study.

Having 250 women as the sample size was also expected to provide a reasonable number of cases of each of the more common target conditions from which to estimate the sensitivity of MRI for diagnosis. We anticipated a high sensitivity of MRI for detecting common pathological causes of CPP; therefore, we based our calculations on an anticipated sensitivity of 80% for any particular condition (a sensitivity of 90% has also been provided for comparison; *Table 4*). We then computed the 95% confidence intervals (CIs) for these sensitivities for a range of prevalences of any particular condition (see *Table 4*). These figures could equally apply to specificity. For a target condition with a prevalence of 30% or more, we expected to be able to reliably rule out a sensitivity or specificity of < 70% if the 'true' sensitivity or specificity was > 80%.

In May 2013, when the target sample size had already been slightly exceeded (269 women), the trial team, together with the input of the independent oversight committees, decided to extend recruitment until the end of September 2013, with a revised target of 340 women, and the protocol was amended accordingly. At this stage, the study was recruiting faster than was originally planned. The aim of extending recruitment was to improve the statistical precision of the accuracy estimates for the less prevalent gynaecological

TABLE 4 Range of 95% CIs for varying prevalences

Assumed sample size of 250 women		Assumed sensitivity (%)	
Prevalence (%)	Number of cases	80	90
50	125	72–87	84–95
40	100	71–87	82–95
30	75	69–88	82–96
20	50	66–90	78–97

causes, and this revised plan was ratified by the external Trial Steering Committee. For example, a sensitivity of 90% would have a lower 95% CI boundary of 80% for conditions occurring in 30% of women in a sample size of 250, and in conditions occurring in 20% of women in a sample size of 340. However, between May 2013 and September 2013, recruitment to the study unexpectedly started to slow and this revised target was not met, with a final sample size of 291 participants at the start of the diagnostic study.

Data analysis

Reliability estimates

Agreement between the three individual members of the EIP in their diagnosis (for 'any gynaecological cause' and each gynaecological cause separately) was analysed using the methodology proposed by Fleiss,⁷² which is a generalisation of the Cohen's kappa statistic to the measurement of agreement among multiple raters. In addition, the agreement between the consensus panel diagnosis, made with and without the MRI findings, was analysed using the Cohen's kappa agreement statistic. In both cases, the kappa value was presented alongside its 95% CI. We also reported the number of cases in which disagreements occurred.

Accuracy estimates

By comparing the index test with the reference standard, the results of the index test were categorised as a true positive, a false positive, a true negative or a false negative. The sensitivity of each test was computed as the proportion of all positive results on the reference standard that were true positives, and the specificity was computed as the proportion of all negative results on the reference standard that were true negatives.

The prevalence of all gynaecological and non-gynaecological conditions in relation to whether or not they were believed to be causing pain (as determined by the reference standard established by the consensus of the EIP) was calculated, along with the 95% CIs, using binomial exact methods.⁸⁶ For gynaecological conditions, estimates of prevalence for whether or not the condition was also observed at laparoscopy were also produced.

The initial analysis looked at the diagnostic accuracy of MRI in being able to detect if a condition actually existed (regardless of whether or not it was judged to be the cause of pain), using the laparoscopy report as the reference standard; MRI data were taken directly from the MRI scan report using the set of rules described in *Table 2*. Accuracy estimates for this additional analysis (sensitivity and specificity, along with 95% CIs) were calculated, and Fisher's exact test was used to explore if MRI and laparoscopy detected the same women with a presence of each target condition.

The primary analysis involved calculations of sensitivity and specificity (and the associated 95% CIs) for the presence or absence of each structural gynaecological cause of pain; data were taken from the binary yes/no responses of the reference standard before and after the MRI data were revealed. Receiver operating characteristic (ROC) curves were also constructed for each test for each condition, using the certainty estimates (0 = 'no chance/almost no chance' to 10 = 'certain, practically certain') provided by the clinician and the binary responses of the reference standard. The area under the ROC (AUROC) curve and the 95% CI were estimated for each test. Differences between the AUROC estimates of each of the four diagnoses were then calculated using a non-parametric approach for correlated data,⁷¹ and these differences are presented along with their 95% CIs. The same analysis was done to compare the responses after the MRI data were revealed (from the independent gynaecologist) with the responses after the laparoscopy had been undertaken (from the treating gynaecologist), to compare the accuracy of MRI-based diagnoses with laparoscopy-based diagnoses. Only participants with laparoscopic, MRI and EIP diagnoses were included in the analysis.

The proportion of women for whom a diagnostic and/or therapeutic laparoscopy could be avoided was calculated by adding up the number of women needing treatment other than laparoscopy (women with adenomyosis, fibroids or PID) and the number of women not needing any treatment (as a result of the absence of any gynaecological cause), which were correctly identified from the MRI scan.

All analyses were performed using SAS, version 9.4 (SAS Institute Inc., Cary, NC, USA).

Chapter 3 Diagnostic study results

Recruitment

The recruitment of participants started in December 2011 and closed in September 2013. A total of 1667 women were screened. A total of 435 women, who consented to participate, were recruited into the study from 26 centres (see *Appendix 1, Table 29*). Ultimately, 287 women contributed to the analysis (*Figure 3*).

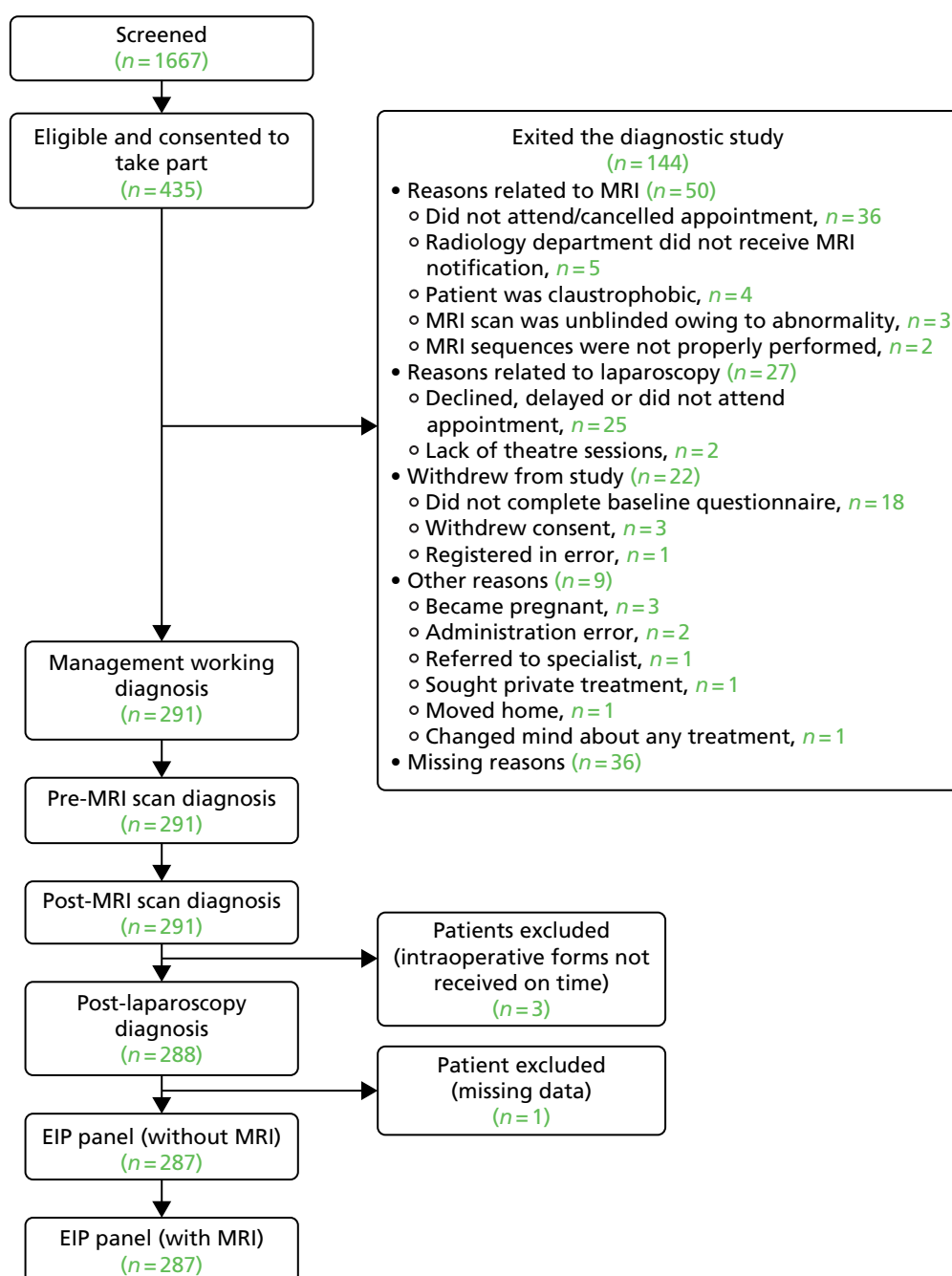


FIGURE 3 Flow chart of participants, showing the numbers who provided data at each stage.

The original sample size of 250 participants was reached ahead of schedule, so the sample size was revised to 340 participants. The sample size was increased to improve the precision of estimates for all target conditions, but most notably for the less common target conditions.

Characteristics of participants

The mean age of the women in our sample was 31.6 years [standard deviation (SD) 8.3 years]. Of these, 197 women (68.5%) were in employment and 96 (33.5%) had received a university-level education. There were 216 white women (74.5%). The duration of pain was, on average, 4.2 years (SD 4.8 years); pain was present on at least 1 day every week in 226 women (79.5%), with a mean pain score of 7.1 points (SD 2.8 points) during menstrual periods and 6.4 points (SD 2.6 points) before menstrual periods. The average age at menarche was 12.8 years (SD 1.7 years). Heavy periods were a feature in 147 women (58.6%). Around three-quarters of women ($n = 212$; 73.4%) were sexually active, with pain during intercourse featuring over the last 20.3 months (SD 20.7 months). Contraceptive hormones had been prescribed for pain to 162 women (55.7%). Chlamydia testing was undertaken in 202 women (70.4%), of whom 176 (89.3%) had received negative results. Laparoscopy and MRI had been used for diagnosis in the past in 80 (27.5%) and 30 women (10.3%), respectively. Infertility was not a complaint in 206 women (85.1%). Further details of participant characteristics are shown in *Table 5* and *Appendix 1, Tables 30 and 31*.

The median duration between the MRI scan and the diagnostic laparoscopy was 26 days (interquartile range of 9–54 days).

Reliability of the expert independent panel diagnoses

We investigated the reliability of the EIP reference diagnoses by looking first at the consensus ratings made with and without access to the MRI report and, second, at the agreement between individual raters in the independent expert group.

Table 6 reports the prevalence of each condition as defined by the EIP with and without access to the MRI report. We considered the primary reference standard to be the EIP consensus statement without the MRI report, as this would reduce the risk of incorporation bias in assessing the accuracy of the MRI report. However, we were interested in noting when the reference standard consensus diagnoses differed, as it was known that an MRI scan may be the only test to identify several of the conditions in some women.

Table 6 shows that the difference in the number of cases detected was between zero and two for all conditions other than adenomyosis (difference of five cases detected), and cases with psychological and psychosexual causes (difference of three cases detected). Although there are clinical reasons why adenomyosis can only be detected by a MRI scan, which are likely to explain the difference, there are no reasons to explain the difference for cases with psychological and psychosexual causes. We therefore report comparisons with the second-stage MRI scan-informed EIP consensus reference diagnosis for analysis of adenomyosis, but not for any other conditions.

To assess the reliability of the EIP diagnoses, we considered the agreement between the three individual ratings made by the panel members prior to any group discussion. We computed the kappa statistic across the three raters to describe the agreement beyond that explained by chance. Kappa values above 0.80 were considered to indicate excellent agreement, whereas those between 0.60 and 0.79 indicate good agreement and those below 0.60 indicate poor or moderate agreement.

TABLE 5 Characteristics of participants

Characteristic	Value
Age (years), mean (SD, <i>n</i>)	31.6 (8.3, 291)
Marital status, ^a <i>n</i> (%)	
Single	96 (33.2)
Living together	66 (22.8)
Married	110 (38.1)
Separated/divorced	17 (5.9)
Employment, ^b <i>n</i> (%)	
Full-time employment	128 (44.4)
Part-time employment	50 (17.4)
Self-employed	19 (6.6)
Caring for children	35 (12.2)
Student	22 (7.6)
Unemployed	34 (11.8)
Highest level of education achieved, ^c <i>n</i> (%)	
No qualifications	17 (5.9)
GCSE/O level/NVQ1–2	94 (32.9)
A level/BTEC qualification/NVQ3–4	79 (27.6)
University degree	67 (23.4)
Postgraduate degree	29 (10.1)
Ethnic group, ^d <i>n</i> (%)	
Asian/Asian British	33 (11.3)
Black/black British	18 (6.2)
White	216 (74.5)
Mixed	12 (4.1)
Any other ethnic group	9 (3.1)
Do not wish to say	2 (0.7)

A level, Advanced level; BTEC, Business and Technology Education Council; GCSE, General Certificate of Secondary Education; NVQ, National Vocational Qualification; O level, Ordinary level.

a Two missing values.

b Three missing values.

c Five missing values.

d One missing value.

The panel were in excellent agreement in identifying deep-infiltrating endometriosis and endometrioma of the ovary as causes of pelvic pain, and in good agreement in identifying superficial peritoneal endometriosis and adenomyosis when the MRI reports were used in the reference standard. Lower levels of agreement were noted for deciding that adhesions, fibroids and PID were the cause of pelvic pain. The expert panel failed to demonstrate adequate reliability in determining whether any of the non-gynaecological causes were the cause of pelvic pain for further detailed examination of these causes.

TABLE 6 Reliability of the EIP reference diagnoses

Cause of pelvic pain	Prevalence (<i>N</i> = 287), <i>n</i> (%)		Reliability	
	Reference diagnosis without MRI findings	Reference diagnosis with MRI findings	Assessors' kappa ^a (95% CI)	Number of cases disputed ^b
Idiopathic for gynaecological conditions	156 (54.4)	153 (53.3)	0.75 (0.69 to 0.84)	40
Superficial peritoneal endometriosis	71 (24.7)	71 (24.7)	0.74 (0.67 to 0.81)	32
Deep-infiltrating endometriosis	40 (13.9)	42 (14.6)	0.96 (0.89 to 0.99)	3
Endometrioma of the ovary	11 (3.8)	13 (4.5)	0.80 (0.73 to 0.87)	6
Adhesions	24 (8.4)	25 (8.7)	0.54 (0.47 to 0.61)	35
Ovarian cysts	1 (0.3)	0 (0)	–	0
Adenomyosis	11 (3.8)	16 (5.6)	0.32 (0.26 to 0.40)	24
			0.60 (0.53 to 0.67)#	22#
Fibroids	4 (1.4)	5 (1.7)	0.53 (0.46 to 0.60)	7
PID	10 (3.5)	10 (3.5)	0.52 (0.45 to 0.69)	10
Urinary causes	18 (6.3)	16 (5.6)	0.35 (0.28 to 0.42)	35
Musculoskeletal causes	48 (16.7)	47 (16.4)	0.51 (0.44 to 0.58)	39
Gastrointestinal causes	24 (8.4)	23 (8.0)	0.51 (0.44 to 0.58)	34
Psychological/psychosexual causes	78 (27.2)	75 (26.1)	0.53 (0.46 to 0.59)	59

a Kappa agreement (95% CI) between the three EIP members. Kappa for agreement between the three assessors for the reference diagnosis made without MRI (stage 1 – see *Figure 1*) unless indicated # where reference diagnosis including MRI was used (stage 2 – see *Figure 1*).

b Number of cases where at least two EIP members disagreed with each other.

Prevalence of conditions being the cause of pelvic pain or observed on magnetic resonance imaging scans or through laparoscopy

Table 7 reports the prevalence of the causes of pelvic pain as judged by the EIP and observed through laparoscopy and MRI scans. For many conditions (superficial peritoneal endometriosis, endometrioma of the ovary, adhesions, ovarian cysts, adenomyosis and fibroids), the prevalence of observing the condition through MRI scans and/or laparoscopy was higher than the prevalence according to the judgements of the EIP, reflecting that, in many women, the conditions may be observed but not judged to be the cause of pelvic pain.

Superficial peritoneal endometriosis was identified as the most common gynaecological cause of pelvic pain in 25% of women, followed by deep-infiltrating endometriosis in 14%, adhesions in 8%, endometrioma of the ovary, adenomyosis and PID each in 4% and fibroids and ovarian cysts in ≤ 1%. Of the non-gynaecological causes, psychological/psychosexual causes were the most common in 27% of women, musculoskeletal causes in 17%, gastrointestinal causes in 8% and urinary causes in 6%. Fifty-four per cent of women were judged to have idiopathic pelvic pain with no structural gynaecological cause identified.

The three endometriosis conditions (superficial peritoneal, deep-infiltrating and endometrioma of the ovary) and ovarian cysts were observed twice as often with laparoscopy, and adhesions three times as often, as with MRI scans. Adenomyosis, fibroids and PID were noted more often through MRI than through laparoscopy. No MRI diagnostic criteria were available for the non-gynaecological conditions.

TABLE 7 Prevalence of conditions being the cause of pelvic pain or observed through MRI scans or laparoscopy

Cause of pelvic pain	Judged as being the cause of pelvic pain, n (%)		Observation by technique, n (%)	
	Reference diagnosis without MRI	Reference diagnosis with MRI	Laparoscopy (N = 287)	MRI ^a (N = 287)
Idiopathic (no structural gynaecological cause)	156 (54.4)	153 (53.3)	75 (26.1)	49 (17.1) ^b 148 (51.6) ^c
Superficial peritoneal endometriosis	71 (24.7)	71 (24.7)	120 (41.8)	No criteria
Deep-infiltrating endometriosis	40 (13.9)	42 (14.6)	37 (12.9)	3 (1.0)
Endometrioma of the ovary	11 (3.8)	13 (4.5)	68 (23.7)	31 (10.8)
Adhesions	24 (8.4)	25 (8.7)	109 (39.6)	37 (12.9)
Ovarian cysts	1 (0.3)	0 (0)	56 (19.7)	27 (9.4)
Adenomyosis	11 (3.8)	16 (5.6)	21 (7.3)	193 (67.2) ^b 34 (11.8) ^c
Fibroids	4 (1.4)	5 (1.7)	24 (8.4)	56 (19.5)
PID	10 (3.5)	10 (3.5)	10 (3.5)	15 (5.2)
Urinary causes	18 (6.3)	16 (5.6)	Not assessed	No criteria
Musculoskeletal causes	48 (16.7)	47 (16.4)	Not assessed	No criteria
Gastrointestinal causes	24 (8.4)	23 (8.0)	Not assessed	No criteria
Psychological/psychosexual causes	78 (27.2)	75 (26.1)	Not assessed	No criteria

^a See Table 2 for definitions used for MRI findings.
^b Values when MRI Table 2 definition 1 of adenomyosis is applied.
^c Values when MRI Table 2 definition 2 of adenomyosis is applied.

Figure 4 shows the relationships between the causes of pelvic pain based on the EIP diagnoses without the MRI report for structural gynaecological causes. Superficial peritoneal endometriosis, deep-infiltrating endometriosis and endometrioma of the ovary often occurred together: 17 out of 40 women with deep-infiltrating endometriosis also had superficial peritoneal endometriosis, 7 out of 11 women with endometrioma of the ovary also had deep-infiltrating endometriosis and 7 out of 11 women with endometrioma of the ovary also had superficial peritoneal endometriosis. Adhesions also appeared in combination with the three types of endometriosis.

Connecting lines indicate the number of women with both diagnoses. A total of 131 women (46%) had structural gynaecological causes and are included in Figure 4.

Comparison of observations on magnetic resonance imaging scans against a reference standard of observations made through laparoscopy

If MRI were to be useful at identifying a condition as being the cause of pelvic pain, it is a prerequisite that the condition has to be observed in the MRI image. Thus, before investigating the accuracy of MRI for identification of each condition as a cause, we assessed the accuracy of MRI for observing each condition, using a reference standard of the condition being observed through laparoscopy. Although this reference standard was thought to be suitable for identifying the three endometriosis conditions, ovarian cysts and adhesions, it was unsuitable for use as a reference standard for adenomyosis and fibroids, as it was likely to miss many cases that could not be identified from inspection of the outside of the uterus. The estimates of test accuracy for adenomyosis and fibroids should therefore be considered alongside the likelihood of

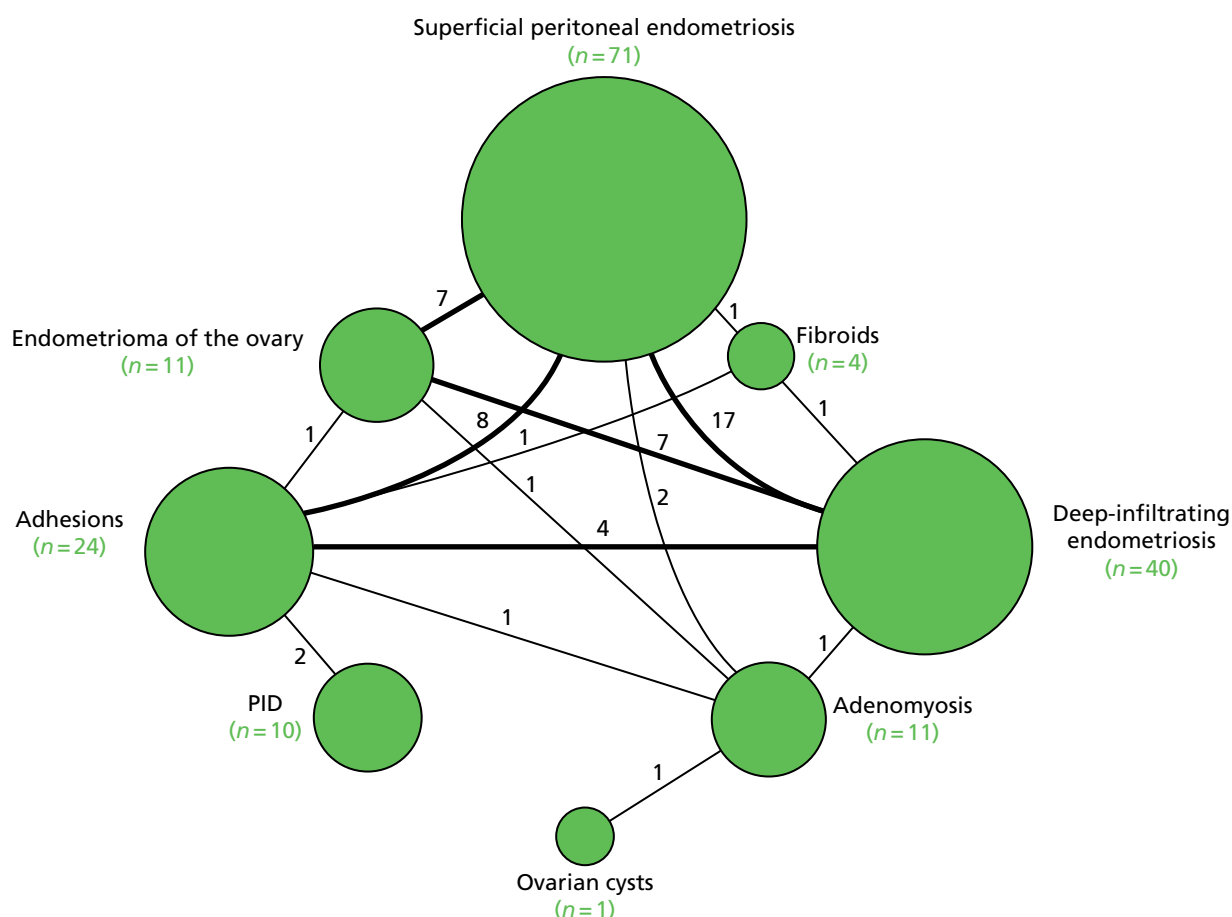


FIGURE 4 Relationships between the structural gynaecological causes.

misclassification of cases on the reference standard, as apparent false positives and false negatives could be caused by misclassification errors in the reference standard, as well as misclassification errors in the index test. The same concern applied to interpretation of MRI scans judged to be idiopathic, as this was defined as being the absence of all conditions, including fibroids and adenomyosis. In addition, no criteria for identifying superficial peritoneal endometriosis or the non-gynaecological conditions on MRI reports were available, and PID was excluded because few cases were detected.

Table 8 shows that MRI had poor sensitivity for detecting many conditions. MRI detected only 1 of the 36 cases of deep-infiltrating endometriosis that had been observed through laparoscopy (with a sensitivity of 3%), and picked up 20 of the 107 cases of adhesions (with a sensitivity of 19%), 11 of the 56 cases of ovarian cysts (with a sensitivity of 20%) and 22 of the 66 cases of endometrioma of the ovary (with a sensitivity of 33%). The number of false positives (reported on the MRI scan but not seen on laparoscopy) for these conditions were low, with specificities between 93% and 99%.

We considered two different MRI definitions of adenomyosis (as defined in Table 2), the first of which was met by 193 women (69%) and the second by 34 women (12%). MRI did not show any relationship with the findings on laparoscopy for either definition of adenomyosis ($p = 0.6$ and $p = 0.9$, respectively). It was not possible to ascertain if this was attributable to the poor ability of laparoscopy or MRI to detect this condition. Observing fibroids on MRI scans was closely linked with observing fibroids on laparoscopy with a sensitivity of 88%, but 35 women had fibroids noted on MRI scans that were not observed on laparoscopy, which could be explained by the inability of laparoscopy to observe fibroids inside the uterus.

TABLE 8 Comparison of observations on MRI scans against a reference standard of observations made through laparoscopy

Cause of pelvic pain	MRI finding	Laparoscopy finding, <i>n</i>		Sensitivity (95% CI)	Specificity (95% CI)	Likelihood ratio (95% CI)		Diagnostic odds ratio (95% CI)	Test of association
		Yes	No			Positive	Negative		
Idiopathic ^a	Yes	16	33	21.3% (12.7% to 32.3%)	84.4% (78.8% to 89.0%)	1.4 (0.8 to 2.3)	0.9 (0.8 to 1.1)	1.5 (0.8 to 2.8)	<i>p</i> = 0.3
	No	59	179						
Idiopathic ^b	Yes	55	93	73.3% (61.9% to 82.9%)	56.1% (49.2% to 62.9%)	1.7 (1.4 to 2.1)	0.5 (0.3 to 0.7)	3.5 (2.0 to 6.3)	<i>p</i> < 0.0001
	No	20	119						
Superficial peritoneal endometriosis	Yes	0	0	0%	100%	–	–	–	–
	No	120	167						
Deep-infiltrating endometriosis	Yes	1	2	2.8% (0.1% to 14.5%)	99.2% (97.1% to 99.9%)	3.5 (0.3 to 37.2)	0.98 (0.93 to 1.04)	3.5 (0.3 to 39.9)	<i>p</i> = 0.33
	No	35	247						
Endometrioma of the ovary	Yes	22	9	33.3% (22.2% to 46.0%)	95.8% (92.2% to 98.1%)	8.0 (3.9 to 16.5)	0.7 (0.6 to 0.8)	11.5 (5.0 to 26.7)	<i>p</i> < 0.0001
	No	44	207						
Adhesions	Yes	20	11	18.7% (11.8% to 27.4%)	93.3% (88.4% to 96.6%)	2.8 (1.4 to 5.6)	0.9 (0.8 to 1.0)	3.2 (1.5 to 6.9)	<i>p</i> = 0.003
	No	87	154						
Ovarian cysts	Yes	11	16	19.6% (10.2% to 32.4%)	92.9% (88.7% to 95.9%)	2.8 (1.4 to 5.6)	0.9 (0.8 to 1.0)	3.2 (2.2 to 4.5)	<i>p</i> = 0.009
	No	45	208						
Adenomyosis ^a	Yes	15	175	75.0% (51.0% to 91.0%)	31.9% (26.0% to 38.0%)	1.1 (0.8 to 1.4)	0.8 (0.4 to 1.7)	1.4 (0.5 to 4.0)	<i>p</i> = 0.6
	No	5	82						
Adenomyosis ^b	Yes	2	31	10.0% (1.2% to 31.7%)	87.8% (83.1% to 91.5%)	0.8 (0.2 to 3.2)	1.0 (0.9 to 1.2)	0.8 (0.2 to 3.6)	<i>p</i> = 0.9
	No	18	222						
Fibroids	Yes	21	35	87.5% (67.6% to 97.3%)	86.4% (81.6% to 90.4%)	6.5 (4.6 to 9.1)	0.1 (0.05 to 0.4)	44.6 (12.3 to 157.4)	<i>p</i> < 0.0001
	No	3	223						
<p>^a Values when MRI <i>Table 2</i> definition 1 of adenomyosis is applied.</p> <p>^b Values when MRI <i>Table 2</i> definition 2 of adenomyosis is applied.</p>									

We defined an idiopathic MRI scan to be one in which no criteria for any of the above gynaecological conditions were met. As we considered two alternative MRI definitions of adenomyosis, there were two corresponding definitions of 'idiopathic'. The first MRI definition categorised the majority of women as having adenomyosis, with only 49 women (18%) judged to have an idiopathic MRI report, and this finding showed no relationship with observing no structural gynaecological cause on laparoscopy ($p = 0.3$; see *Appendix 2, Table 32*).

Using the second definition, 148 women (52%) were categorised as having an idiopathic MRI report, of whom 93 had findings on laparoscopy. Twenty women, defined as idiopathic on laparoscopy, were noted to meet the criteria for a gynaecological condition using MRI. Overall, the idiopathic diagnoses made through laparoscopy and MRI were in agreement in only 61% of women (see *Appendix 2, Table 33*).

Comparison of observations made using magnetic resonance imaging against the expert panel reference standard for the cause of pelvic pain

In a further analysis, we considered whether or not MRI findings were diagnostic of the cause of pelvic pain as stated by the EIP. We compared with the EIP consensus diagnosis based on information excluding MRI for all conditions other than adenomyosis where consensus diagnoses including information from MRI was used (*Table 9*). Most conditions on MRI reports were less likely to be regarded as a cause of pelvic pain than conditions observed on laparoscopy; using MRI reports, superficial peritoneal endometriosis was rated as a cause of pain in 71 women, whereas it was observed in 120 women on laparoscopy; endometrioma of the ovary was noted as a cause in 11 women, but observed in 66 women; adhesions was a cause in 24 women, while noted in 107 women; ovarian cysts was a cause in only 1 woman, but noted in 56 women; adenomyosis was a cause in 16 women, but noted in 20 women; and fibroids was a cause in 4 women, while noted in 24 women. The only exceptions were deep-infiltrating endometriosis, which was regarded as a cause of pain in 39 women, but only noted in 36 women, and idiopathic (lack of a gynaecological structural cause), which was concluded in only 75 women on laparoscopy, but was the conclusion of the EIP in 156 women.

The accuracy of MRI to identify the cause of pelvic pain was inestimable for superficial peritoneal endometriosis and non-gynaecological causes, as no MRI definitions were available. Estimates of test sensitivity for ovarian cysts and fibroids were based on only one and four cases, respectively, and thus were very imprecise.

The sensitivity of MRI for detecting deep-infiltrating endometriosis as a cause of pelvic pain was very low (3%), but close to 80% for endometrioma of the ovary (albeit with wide CIs). The sensitivity for adhesions and adenomyosis using the second definition was between 40% and 50%. For all these conditions, specificity was around 90%.

There was no significant relationship between the MRI findings and the expert panel findings drawing an idiopathic conclusion for the cause of pelvic pain for either definition ($p = 0.9$ and $p = 0.08$, respectively). For the second definition (using the more stringent definition of adenomyosis), 159 out of 287 MRI findings and expert consensus classifications of whether the chronic pain was idiopathic or structural (55%) were in agreement.

The impact of magnetic resonance imaging on confidence and accuracy in making diagnoses

The independent gynaecologist rated the likelihood of each cause of CPP first when provided with all baseline data (including ultrasound reports) and, second, when additionally provided with the report from the MRI scan. This allows an assessment to be made of the degree to which the MRI scan provided additional diagnostic information over and above that available from the baseline assessments. The independent gynaecologist was not provided with the MRI criteria stated in *Table 2*, but was provided with only the measurements made on MRI and the textual report. Each independent gynaecologist rated their certainty that each condition was causing pelvic pain on a scale of 0 (no chance) to 10 (certain). We deduced the certainty that the diagnosis was idiopathic by subtracting the highest score for the gynaecological structural causes from 10. The results for scores of 0–3, 4–6 and 7–10 are presented in *Table 9*, along with the likelihood ratios

TABLE 9 Comparison of observations made using MRI against the expert panel reference standard for the cause of pelvic pain

Cause of pelvic pain	MRI finding	Expert panel consensus, <i>n</i>		Sensitivity (95% CI)	Specificity (95% CI)	Likelihood ratio (95% CI)		Diagnostic odds ratio (95% CI)	Test of association
		Yes	No			Positive	Negative		
Idiopathic ^a	Yes	26	23	16.7% (11.2% to 23.5%)	82.4% (74.8% to 88.5%)	0.9 (0.6 to 1.6)	1.0 (0.9 to 1.1)	0.9 (0.5 to 1.7)	<i>p</i> = 0.9
	No	130	108						
Idiopathic ^b	Yes	88	60	56.4% (48.3% to 64.3%)	54.2% (45.3% to 62.9%)	1.2 (0.9 to 1.5)	0.8 (0.6 to 1.0)	1.5 (0.9 to 2.4)	<i>p</i> = 0.08
	No	68	71						
Superficial peritoneal endometriosis	Yes	–	–	–	–	–	–	–	–
	No	71	216						
Deep-infiltrating endometriosis	Yes	1	2	2.6% (0.1% to 13.5%)	99.2% (97.1% to 99.9%)	3.15 (0.29 to 34.0)	0.98 (0.93 to 1.03)	3.21 (0.28 to 36.30)	<i>p</i> = 0.36
	No	38	244						
Endometrioma of ovary	Yes	9	22	81.8% (48.2% to 97.7%)	91.9% (88.0% to 94.8%)	10.1 (6.2 to 16.4)	0.2 (0.1 to 0.7)	50.9 (10.4 to 250.5)	<i>p</i> < 0.0001
	No	2	249						
Adhesions	Yes	9	28	37.5% (18.8% to 59.4%)	89.2% (84.8% to 92.7%)	3.5 (1.9 to 6.5)	0.7 (0.5 to 0.9)	5.0 (2.0 to 12.4)	<i>p</i> = 0.001
	No	15	232						
Ovarian cysts	Yes	0	27	0% (0% to 97.5%)	90.4% (86.3% to 93.6%)	–	–	–	<i>p</i> = 0.9
	No	1	254						
Adenomyosis ^a [post MRI EIP ^c]	Yes	9 (15)	184 (178)	81.8% (48.2% to 97.7%)	31.6% (26.1% to 37.5%)	1.2 (0.9 to 1.6)	0.6 (0.2 to 2.0)	2.0 (0.4 to 9.8)	<i>p</i> = 0.5
	No	2 (1)	85 (86)	[93.8% (70.0% to 99.8%)]	[32.6% (27.0% to 38.6%)]	[1.4 (1.2 to 1.6)]	[0.2 (0.03 to 1.3)]	[7.2 (0.9 to 55.8)]	[<i>p</i> = 0.03]
Adenomyosis ^b [post MRI EIP ^c]	Yes	0 (7)	34 (27)	0% (0% to 30.8)	87.2% (82.6% to 91.0%)	0 (–)	1.15 (1.1 to 1.2)	0 (–)	<i>p</i> = 0.6
	No	10 (8)	232 (234)	[46.7% (21.3% to 73.4%)]	[89.7% (85.3% to 93.1%)]	[4.5 (2.4 to 8.6)]	[0.6 (0.4 to 0.9)]	[7.6 (2.6 to 22.6)]	[<i>p</i> = 0.001]
Fibroids	Yes	4	52	100% (39.8 to 100)	81.4% (76.4 to 85.8)	5.4 (4.2 to 6.9)	0 (–)	– (–)	<i>p</i> = 0.001
	No	0	228						

^a Values when MRI *Table 2* definition 1 of adenomyosis is applied.

^b Values when MRI *Table 2* definition 2 of adenomyosis is applied.

^c Results for adenomyosis are compared first with the reference panel consensus without MRI findings and, second, with the reference panel consensus diagnosis with MRI findings.

computed to indicate the degree to which high (or low) scores rule each condition in (or out). We measured the diagnostic accuracy across all scores by plotting ROC curves for each condition (*Figure 5*) and summarised the accuracy as the AUROC value. The likelihood ratio and AUROC values were computed for assessments made on baseline evidence alone, and then made with the additional provision of the MRI report; the incremental diagnostic value of the MRI scan is summarised using the change in the AUROC value and by comparing the likelihood ratio values.

An AUROC value of 0.5 would be obtained if diagnoses were made by chance, whereas a test with perfect discrimination would have a value of 1.0. The baseline values of the AUROC were not significantly different from 0.5 for all conditions except adhesions (AUROC value of 0.64), indicating that the independent gynaecologist was not significantly better than chance in producing a diagnosis that agreed with that of the EIP, and only had a weak relationship for adhesions. Provision of the MRI report improved the accuracy of all diagnoses, but the increase in AUROC value was only statistically significant for deep-infiltrating endometriosis

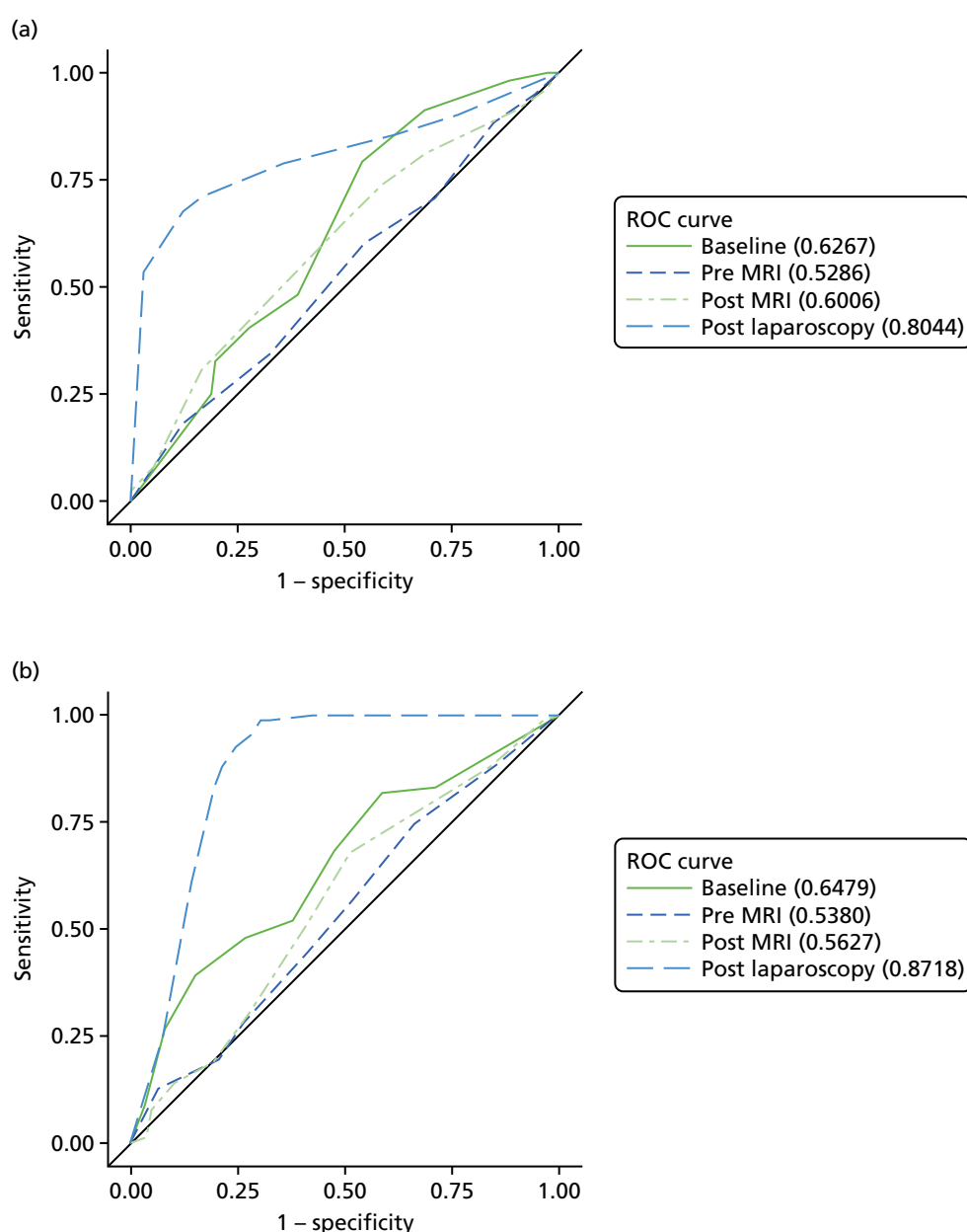


FIGURE 5 Accuracy of assessments of the cause of pelvic pain pre and post MRI and laparoscopy: ROC curves for comparison. (a) Idiopathic; (b) superficial peritoneal endometriosis; (c) deep-infiltrating endometriosis; (d) endometrioma of the ovary; (e) adhesions; and (f) adenomyosis (reference standard with MRI). (*continued*)

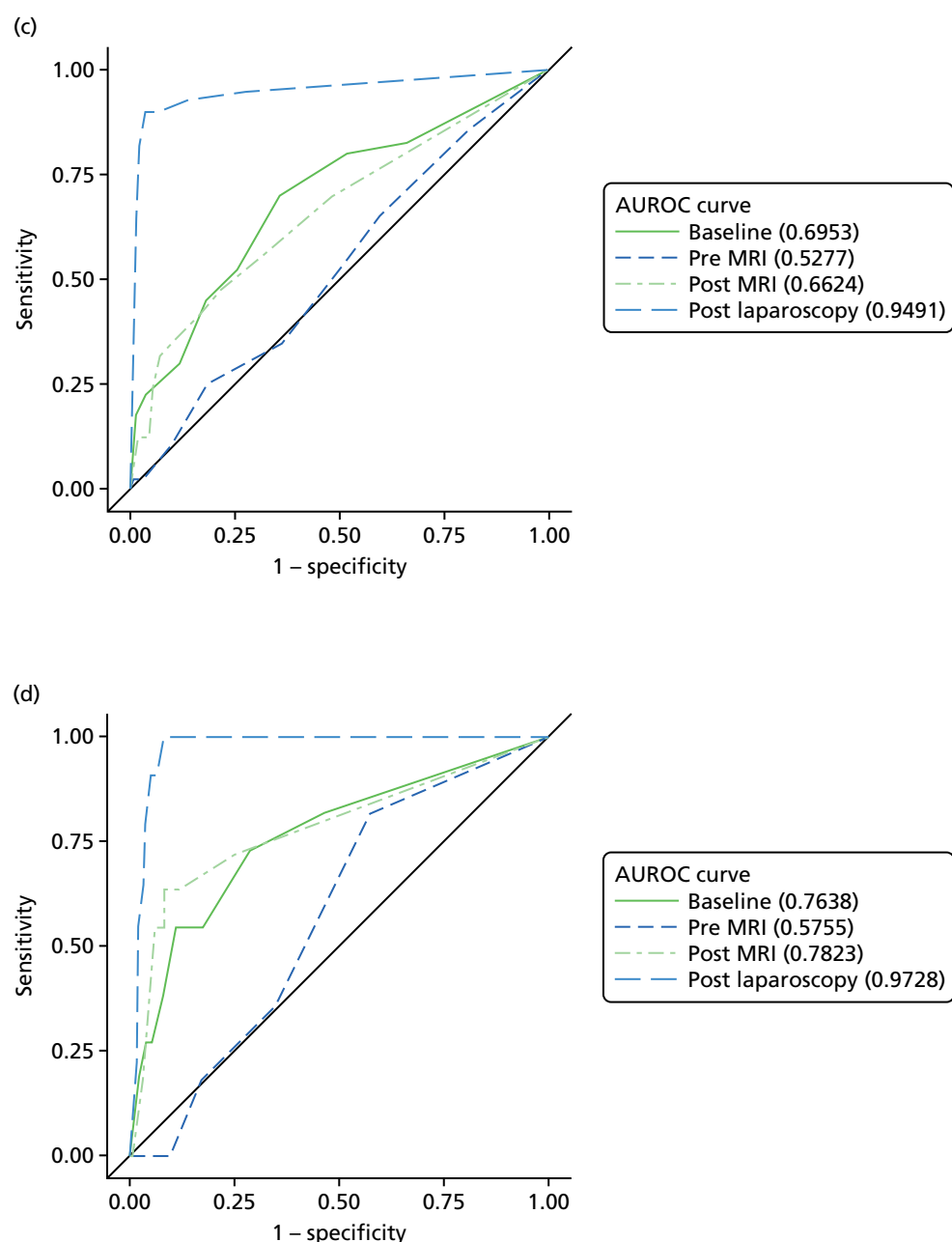


FIGURE 5 Accuracy of assessments of the cause of pelvic pain pre and post MRI and laparoscopy: ROC curves for comparison. (a) Idiopathic; (b) superficial peritoneal endometriosis; (c) deep-infiltrating endometriosis; (d) endometrioma of the ovary; (e) adhesions; and (f) adenomyosis (reference standard with MRI). (*continued*)

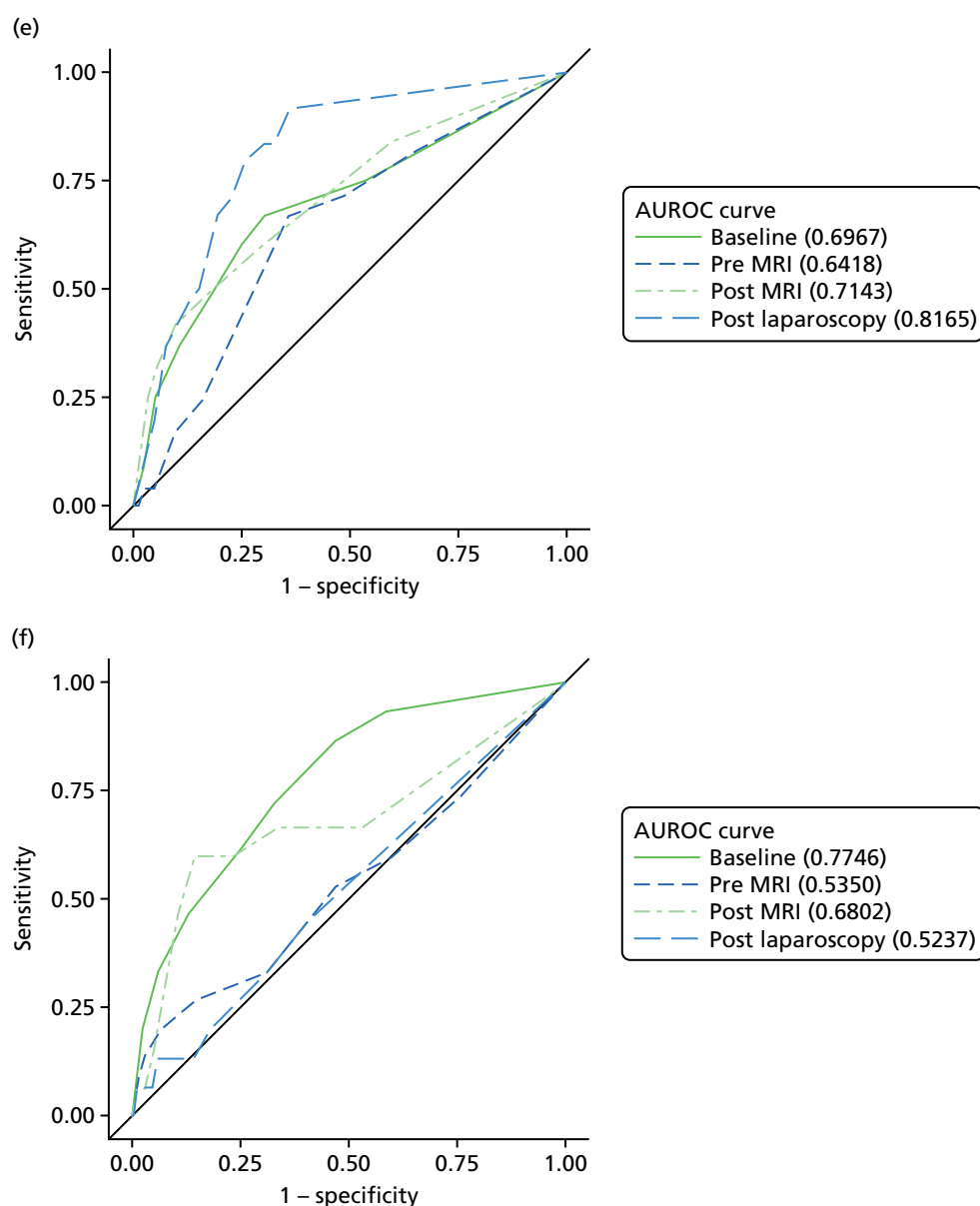


FIGURE 5 Accuracy of assessments of the cause of pelvic pain pre and post MRI and laparoscopy: ROC curves for comparison. (a) Idiopathic; (b) superficial peritoneal endometriosis; (c) deep-infiltrating endometriosis; (d) endometrioma of the ovary; (e) adhesions; and (f) adenomyosis (reference standard with MRI).

($p = 0.006$) and endometrioma of the ovary ($p = 0.02$). Likelihood ratios for scores in the 4–6 and 7–10 categories deep-infiltrating endometriosis and endometrioma of the ovary were around 4–7, which are not high enough to convincingly rule in these diagnoses. Although scores of 7–10 for adhesions had the highest likelihood ratio, at 7.3, they were made in only 15 women, and detected in only 30% of those with adhesions as a cause of pelvic pain (Table 10).

Accuracy of assessments of the cause of pelvic pain pre and post MRI and laparoscopy

Table 11 allows comparisons to be made between the findings from MRI and laparoscopy. As for assessment of the incremental diagnostic value of MRI, we report values for the AUROC, the change in AUROC and the likelihood ratios to describe diagnostic accuracy.

TABLE 10 Impact of MRI on confidence in, and accuracy of, diagnoses made

Cause of pelvic pain		Diagnostic confidence score	Rating							
			Before MRI results revealed				After MRI results revealed			
			Expert panel consensus, <i>n</i>		Likelihood ratio (95% CI)	AUROC (95% CI)	Expert panel consensus, <i>n</i>		Likelihood ratio (95% CI)	AUROC (95% CI)
Yes	No	Yes	No							
Idiopathic	7–10	53	43	1.0 (0.7 to 1.4)	0.53 (0.46 to 0.60)	77	44	1.5 (1.1 to 2.0)	0.60 (0.53 to 0.67)	+0.07 (–0.01 to 0.15; <i>p</i> = 0.06)
	4–6	86	69	1.0 (0.8 to 1.3)		49	47	0.9 (0.6 to 1.2)		
	0–3	17	19	0.8 (0.4 to 1.4)		30	40	0.6 (0.4 to 1.0)		
Superficial peritoneal endometriosis	7–10	9	13	2.1 (0.9 to 4.7)	0.54 (0.46 to 0.61)	5	10	1.5 (0.5 to 4.3)	0.56 (0.49 to 0.64)	+0.02 (–0.05 to 0.10; <i>p</i> = 0.5)
	4–6	28	92	0.9 (0.7 to 1.3)		19	54	1.1 (0.7 to 1.7)		
	0–3	34	111	0.9 (0.7 to 1.2)		47	152	0.9 (0.8 to 1.1)		
Deep-infiltrating endometriosis	7–10	1	2	3.1 (0.3 to 33.3)	0.53 (0.43 to 0.62)	5	6	5.1 (1.7 to 9.4)	0.66 (0.57 to 0.76)	+0.13 (0.04 to 0.23; <i>p</i> = 0.006)
	4–6	9	42	1.3 (0.7 to 2.5)		8	12	4.1 (1.8 to 9.4)		
	0–3	30	203	0.9 (0.8 to 1.1)		27	229	0.7 (0.6 to 0.9)		
Endometrioma of the ovary	7–10	0	2	0	0.58 (0.44 to 0.71)	2	8	6.3 (1.5 to 26.1)	0.78 (0.62 to 0.95)	+0.20 (0.02 to 0.39; <i>p</i> = 0.02)
	4–6	0	24	0		4	14	7.2 (2.8 to 18.2)		
	0–3	11	250	1.1 (1.0 to 1.2)		5	254	0.5 (0.3 to 0.9)		
Adhesions	7–10	1	5	2.2 (0.3 to 18.0)	0.64 (0.53 to 0.75)	6	9	7.3 (2.8 to 18.8)	0.72 (0.60 to 0.83)	+0.08 (–0.03 to 0.18; <i>p</i> = 0.2)
	4–6	5	39	1.4 (0.6 to 3.2)		6	37	1.8 (0.8 to 3.8)		
	0–3	18	219	0.9 (0.7 to 1.1)		12	217	0.6 (0.4 to 0.9)		
Adenomyosis (post-MRI EIP consensus)	7–10	2	7	4.8 (1.1 to 21.4)	0.53 (0.36 to 0.71)	6	22	4.6 (2.2 to 9.8)	0.68 (0.50 to 0.85)	+0.15 (–0.03 to 0.32; <i>p</i> = 0.1)
	4–6	3	80	0.6 (0.2 to 1.8)		4	41	1.7 (0.7 to 4.0)		
	0–3	11	184	1.0 (0.7 to 1.4)		6	208	0.5 (0.3 to 0.9)		

TABLE 11 Difference between MRI and laparoscopy on confidence in, and accuracy of, diagnoses made

Cause of pelvic pain		Rating after								
		MRI results are revealed				Laparoscopy				Change in AUROC (95% CI; <i>p</i> -value)
		Expert panel consensus, <i>n</i>		Likelihood ratio (95% CI)	AUROC (95% CI)	Expert panel consensus, <i>n</i>		Likelihood ratio (95% CI)	AUROC (95% CI)	
		Yes	No			Yes	No			
Diagnostic confidence score	Yes	No	Likelihood ratio (95% CI)	AUROC (95% CI)	Yes	No	Likelihood ratio (95% CI)	AUROC (95% CI)		
Idiopathic	7–10	77	44	1.5 (1.1 to 2.0)	0.60 (0.53 to 0.67)	81	4	17.6 (6.6 to 46.6)	0.80 (0.75 to 0.86)	+0.20 (0.12 to 0.28; <i>p</i> < 0.0001)
	4–6	49	47	0.9 (0.6 to 1.2)		26	18	1.3 (0.7 to 2.2)		
	0–3	30	40	0.6 (0.4 to 1.0)		44	109	0.4 (0.3 to 0.5)		
Superficial peritoneal endometriosis	7–10	5	10	1.5 (0.5 to 4.3)	0.56 (0.49 to 0.64)	54	38	4.2 (3.1 to 5.7)	0.87 (0.83 to 0.91)	+0.31 (0.22 to 0.39; <i>p</i> < 0.0001)
	4–6	19	54	1.1 (0.7 to 1.7)		12	15	2.4 (1.2 to 4.8)		
	0–3	47	152	0.9 (0.8 to 1.1)		5	156	0.1 (0.04 to 0.2)		
Deep-infiltrating endometriosis	7–10	5	6	5.1 (1.7 to 9.4)	0.66 (0.57 to 0.76)	33	6	33.0 (14.7 to 73.3)	0.95 (0.90 to 0.99)	+0.29 (0.16 to 0.35; <i>p</i> < 0.0001)
	4–6	8	12	4.1 (1.8 to 9.4)		3	5	3.6 (0.9 to 14.4)		
	0–3	27	229	0.7 (0.6 to 0.9)		4	228	0.1 (0.04 to 0.3)		
Endometrioma of ovary	7–10	2	8	6.3 (1.5 to 26.1)	0.78 (0.62 to 0.95)	9	10	21.8 (11.2 to 42.6)	0.97 (0.95 to 0.99)	+0.19 (0.03 to 0.36; <i>p</i> = 0.02)
	4–6	4	14	7.2 (2.8 to 18.2)		1	7	3.5 (0.5 to 25.7)		
	0–3	5	254	0.5 (0.3 to 0.9)		1	250	0.1 (0.01 to 0.6)		
Adhesions	7–10	6	9	7.3 (2.8 to 18.8)	0.72 (0.60 to 0.83)	12	39	3.3 (2.0 to 5.4)	0.82 (0.74 to 0.90)	+0.10 (–0.05 to 0.26; <i>p</i> = 0.2)
	4–6	6	37	1.8 (0.8 to 3.8)		7	28	2.7 (1.3 to 5.4)		
	0–3	12	217	0.6 (0.4 to 0.9)		5	189	0.3 (0.1 to 0.6)		
Adenomyosis (post-MRI EIP consensus)	7–10	6	22	4.6 (2.2 to 9.8)	0.68 (0.50 to 0.85)	1	12	1.5 (0.2 to 10.5)	0.52 (0.38 to 0.67)	–0.15 (–0.33 to 0.02; <i>p</i> = 0.09)
	4–6	4	41	1.7 (0.7 to 4.0)		1	25	0.7 (0.1 to 4.8)		
	0–3	6	208	0.5 (0.3 to 0.9)		13	226	1.0 (0.8 to 1.2)		

The accuracy of diagnoses of the causes of pelvic pain made by laparoscopy was high for superficial peritoneal endometriosis (AUROC = 0.87), deep-infiltrating endometriosis (AUROC = 0.95) and endometrioma of the ovary (AUROC = 0.97). Certainty scores of 7 and above could accurately rule for deep-infiltrating endometriosis [LR (likelihood ratio) = 33] and endometrioma of the ovary (LR = 22), and low scores of 0–3 could rule out all three conditions (LR = 0.1).

Poorer performance was noted for laparoscopy diagnoses of idiopathic pelvic pain (AUROC = 0.80) and of adhesions being the cause of pelvic pain (AUROC = 0.82). However, a high score for idiopathic pelvic pain (computed as 10 minus the highest score for any other condition) provided convincing evidence (LR = 18) to rule in an idiopathic diagnosis. Laparoscopy was no better than chance in diagnosing adenomyosis as the cause of pelvic pain (AUROC = 0.52).

For four diagnoses (idiopathic pelvic pain, superficial peritoneal endometriosis, deep-infiltrating endometriosis and endometrioma of the ovary), laparoscopy was significantly more accurate than MRI-based diagnoses in identifying the cause of pelvic pain.

However, there are two reasons to consider these comparisons as being at risk of bias, and that they may potentially overestimate the magnitude of differences between the tests in favour of laparoscopy. First, *Figure 5* shows that the ROC curves for baseline assessments made by the treating gynaecologist (labelled 'baseline') and the independent gynaecologist (labelled 'pre MRI'), based on the same information obtained from history, examination and ultrasound, are not comparable. For all six conditions presented in *Figure 5*, the treating gynaecologist had superior diagnostic accuracy to the independent gynaecologist, which possibly reflects that the diagnoses of the treating clinician benefited from direct interaction with the patient, which the blinded independent gynaecologist lacked (more detail on idiopathic CPP diagnoses is given in *Appendix 2, Tables 34 and 35*). Given that the post-laparoscopy and post-MRI diagnoses were made by the treating gynaecologist and the independent gynaecologist, respectively, the bias introduced by this is likely to have led to overestimation of the difference.

Second, the EIP had access to the laparoscopy findings when making their reference diagnoses for all conditions, which is likely to have introduced a degree of incorporation bias, leading to overestimation of the accuracy of laparoscopy.

The diagnoses of adhesions showed no significant differences in accuracy between any of the four diagnoses for adhesions. This is likely, in part, to be explained by misclassification in the reference standard, which is suspected through the low agreement rates observed between the assessors (see *Table 6*).

Adenomyosis was the only diagnosis of CPP for which the post MRI test accuracy was superior to the post laparoscopy accuracy, despite the difference not being statistically significant. Intriguingly, the diagnosis of adenomyosis made by the recruiting clinician prior to either test was found to be the most accurate.

Use and recommendations for therapeutic laparoscopy

For the final analyses, women were categorised according to whether or not they would benefit from therapeutic laparoscopy. A test pathway incorporating MRI prior to laparoscopy would be of benefit to women should it correctly identify women who do not require therapeutic laparoscopy, as this is commonly undertaken during the diagnostic laparoscopic procedure.

Table 12 compares the accuracy of diagnoses made post MRI for the key conditions indicating the need for therapeutic laparoscopy, and the recommendations for therapeutic laparoscopy by the independent gynaecologist, following their assessment of the MRI report. Fifty-two out of the 112 women (46%) with the key conditions indicating the need for therapeutic laparoscopy were rated with diagnostic confidence scores of 5 or higher, although the independent gynaecologist rated more women (78/112; 70%) as potentially benefiting from a therapeutic laparoscopy, having seen the MRI report. Women with superficial peritoneal endometriosis were the most poorly identified.

TABLE 12 Use and recommendations for therapeutic laparoscopy

Cause of pelvic pain	Prevalence, % (reference standard)	Therapeutic laparoscopy, <i>n</i> (%)		Diagnostic confidence score post MRI test, <i>n</i> (%)			Therapeutic laparoscopy recommended following MRI test
		Recommended by the EIP	Performed (two missing)	≥ 3	≥ 5	≥ 7	
Conditions for which therapeutic laparoscopy is recommended							
Superficial endometriosis	71 (24.7)	69 (97)	58 (83)	48 (68)	13 (18)	5 (7)	49 (69)
Deep-infiltrating endometriosis	40 (13.9)	38 (95)	31 (78)	14 (35)	9 (23)	5 (13)	30 (75)
Endometrioma of the ovary	11 (3.8)	11 (100)	10 (91)	7 (64)	6 (55)	2 (18)	10 (91)
Any endometriosis	96 (33.5)	92 (96)	78 (82)	70 (73)	35 (36)	16 (17)	67 (70)
Adhesions	24 (8.4)	23 (96)	18 (75)	14 (58)	10 (42)	6 (25)	18 (75)
Any endometriosis, adhesions or ovarian cysts	112 (39.0)	106 (95)	92 (83)	91 (81)	52 (46)	31 (28)	78 (70)
Conditions for which therapeutic laparoscopy is not recommended							
Idiopathic	72 (25.0)	13 (18)	34 (47)	60 (83)	36 (50)	8 (11)	49 (68)
Non-gynaecological cause only	90 (31.0)	7 (8)	36 (40)	81 (90)	41 (46)	13 (14)	59 (66)
Anything other than endometriosis or adhesions or ovarian cysts	175 (61.0)	21 (12)	75 (43)	165 (94)	104 (59)	44 (25)	116 (66)

Of the 175 women with gynaecological conditions for which therapeutic laparoscopy would have no clear benefit, non-gynaecological conditions or no structural cause, the post-MRI diagnosis rated 104 (59%) with confidence scores of 5 or above for these conditions (with lower ratings for any endometriosis or adhesions). However, 116 out of the 175 women (66%) were rated as being likely to benefit from therapeutic laparoscopy, based on the interpretation of the MRI report by the independent gynaecologist.

Table 13 compares the ability of the independent gynaecologist to correctly identify women who would, and women who would not, benefit from therapeutic laparoscopy for three situations: (1) if a woman received an EIP diagnosis of superficial endometriosis, deep-infiltrating endometriosis, endometrioma of the ovary or adhesions; (2) if the EIP directly rated a woman as being likely to benefit from a therapeutic laparoscopy; and (3) if a woman was recorded on the operative sheet as having received a therapeutic component during the study laparoscopy.

For all three comparisons, the MRI assessment identified 70% of women categorised as requiring therapeutic laparoscopy (sensitivity). However, the MRI assessment wrongly suggested that two-thirds of women not meeting these criteria did require therapeutic laparoscopy (specificity).

Inspection of the predictive values provides information on the number of errors in decision-making that would be made should the MRI report be used to inform the decision to progress to laparoscopy. As the three criteria were met by increasing proportions of women [the first criterion was met by 112 women (39%), the second by 127 women (44%) and the third by 167 women (58%)], the predictive values of MRI recommendations changed according to the comparison that was made (see Table 13).

TABLE 13 Independent gynaecologist determination of benefit from therapeutic laparoscopy

Therapeutic laparoscopy recommended following MRI test	EIP-identified condition amenable to therapeutic laparoscopy		Therapeutic laparoscopy recommended by the EIP		Therapeutic laparoscopy performed	
	Yes	No	Yes	No	Yes	No
Yes, <i>n</i>	78	116	88	106	117	76
No, <i>n</i>	34	59	39	54	50	42
Sensitivity (95% CI)	69.6% (60.2% to 78.0%)		69.3% (60.5% to 77.2%)		70.1% (62.5% to 76.9%)	
Specificity (95% CI)	33.7% (26.8% to 41.2%)		33.8% (26.5% to 41.6%)		35.6% (27.0% to 44.9%)	
PPV (95% CI)	40.2% (33.2% to 47.5%)		45.4% (38.2% to 52.6%)		60.6% (53.3% to 67.6%)	
NPV (95% CI)	63.4% (52.8% to 73.1%)		58.1% (47.4% to 68.2%)		45.7% (35.2% to 56.4%)	
NPV, negative predictive value; PPV, positive predictive value.						

First, compared with an EIP recommendation that a woman would benefit from therapeutic laparoscopy, a recommendation based on the MRI findings to proceed to laparoscopy would be correct only for 45% of women: 55% would undergo the procedure unnecessarily. In comparison, the alternative pathway of all women receiving laparoscopy leads to 56% of women receiving the procedure unnecessarily. A recommendation to not proceed to laparoscopy would be correct in 58% of women, whereas 42% of women not receiving laparoscopy would have benefited from it.

Extrapolating from these figures, and assuming the same prevalence of conditions, in a cohort of 1000 women, deciding to proceed to laparoscopy based on MRI findings would lead to 369 women who did not require laparoscopy receiving it, and 136 women who did require laparoscopy not receiving it. In a 'laparoscopy for all' strategy, 557 women who did not require laparoscopy would receive it.

If we compare the recommendation for therapeutic laparoscopy from the MRI findings against the actual diagnoses made by the EIP, the number of women receiving laparoscopy unnecessarily would increase to 404 per 1000 women, and the number of women not receiving laparoscopy who did require it would decrease to 118. If the comparison was made against the actual use of therapeutic laparoscopy, the number of women receiving laparoscopy unnecessarily would decrease to 264 per 1000 women, and the number not receiving laparoscopy who did require it would increase to 174.

Chapter 4 Economic evaluation

Introduction

One of the factors that contributes to the lack of speedy diagnosis of CPP among women is that laparoscopy is perceived to be an invasive, expensive and potentially risky procedure, and that it is used far too frequently in the NHS, as a significant proportion of women have no pathology identified.⁷³ MRI represents a less invasive and, potentially, less expensive procedure than laparoscopy, and as such may be the preferred option for the woman or for the health-care provider. However, as shown in previous chapters of this report, the test accuracy of diagnostic laparoscopy in detecting the structural causes of CPP is greater than that for MRI, meaning that women are more likely to be properly diagnosed and then access the appropriate treatment following a diagnostic laparoscopy than with a MRI scan. In this chapter, all of these factors are considered as part of a wider cost-effectiveness analysis, the objective of which is to examine the cost-effectiveness of MRI compared with diagnostic laparoscopy in women with undiagnosed CPP. However, the study design used in the research meant that all women received all tests, and from the data collected, ROC curves were obtained for both MRI and laparoscopy which could then inform their sensitivity and specificity values based on alternative cut-off points. This meant that the actual sensitivity and specificity values for laparoscopy, as used in usual practice, could not be informed by this study, and so consequently it was not possible to parameterise a usual-practice scenario in this economic analysis. Therefore, we consider MRI and laparoscopy with alternative sensitivity and specificity pairs (as informed by the ROC curves) as separate strategies and examine which test and sensitivity and specificity values is the most cost-effective for women with CPP. For completeness, we also compare these strategies with a no-testing scenario, in which a patient receives neither MRI nor laparoscopy.

Methods

Developing the model structure

The comparison of the pathways undertaken by women in this study is best represented in a modelling framework, in which the alternative pathways that women can follow are compared in order to diagnose the cause of their CPP. As all women in the MEDAL study were tested with MRI scans and diagnostic laparoscopy, a model was required to simulate the outcomes of each of the patient pathways when only one test was administered. A model allows explicit representation of the impact of the accuracy of the tests, the costs to the health service and the impact on the health-related QoL experienced by the women undergoing a particular pathway.

A model was developed via consultation with the research team, drawing on key clinical and modelling expertise. A decision tree was applied using TreeAge Pro 2001 software (TreeAge Software, Inc., Williamstown, MA, USA). A 6-month time horizon was adopted, as this was the period of data collection, and, given that there was very little chance of events recurring on the patient pathways (such as a patient receiving multiple laparoscopies), it was felt that this approach was the most appropriate.

Women entered the model having been identified as having CPP and being eligible for diagnostic laparoscopy. Two different patient pathways were compared, which describe alternative approaches to the diagnosis and treatment of the causes of CPP, a summary of which can be found in *Figure 6*. *Figures 7–10* show the patient pathways considered in this analysis, which together form the model structure for the decision tree used in the cost-effectiveness analysis. For both pathways, a baseline examination was conducted, which comprised a patient history followed by a physical examination, and an ultrasound scan

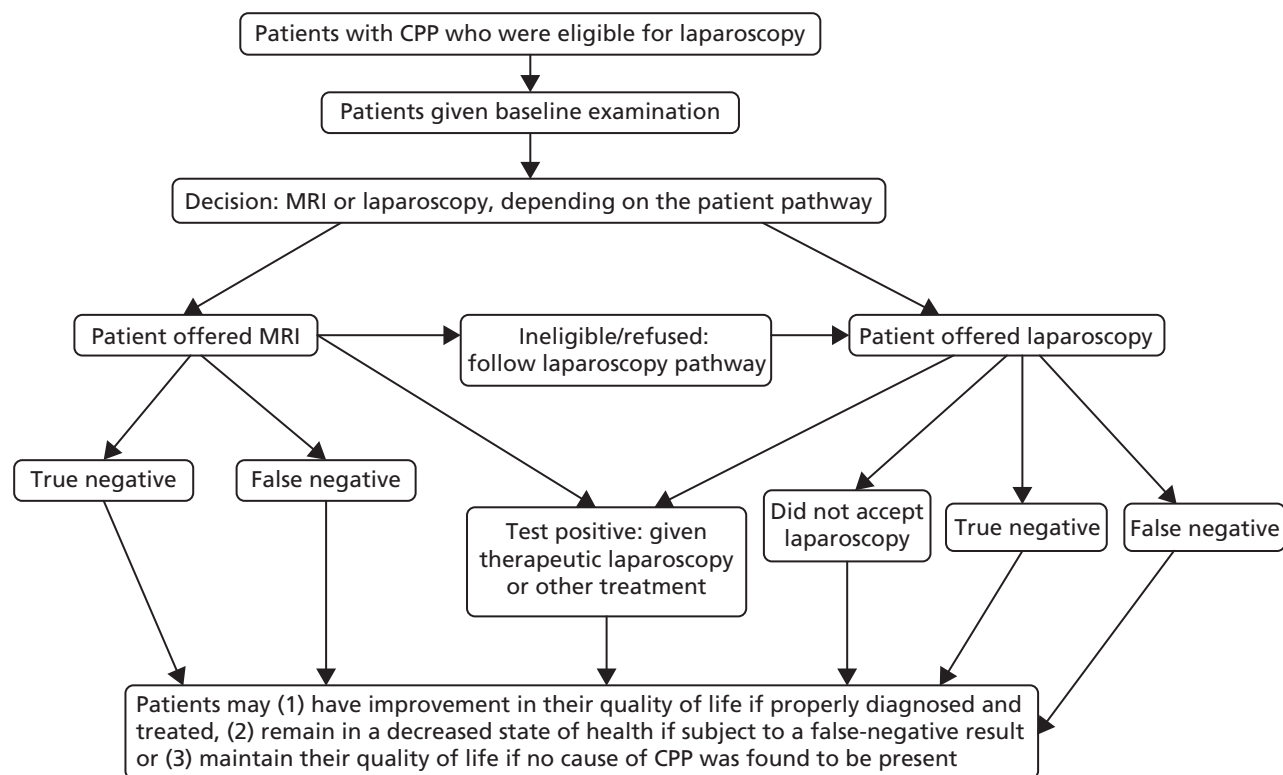


FIGURE 6 Women with CPP who were eligible for laparoscopy, who received either MRI or laparoscopy.

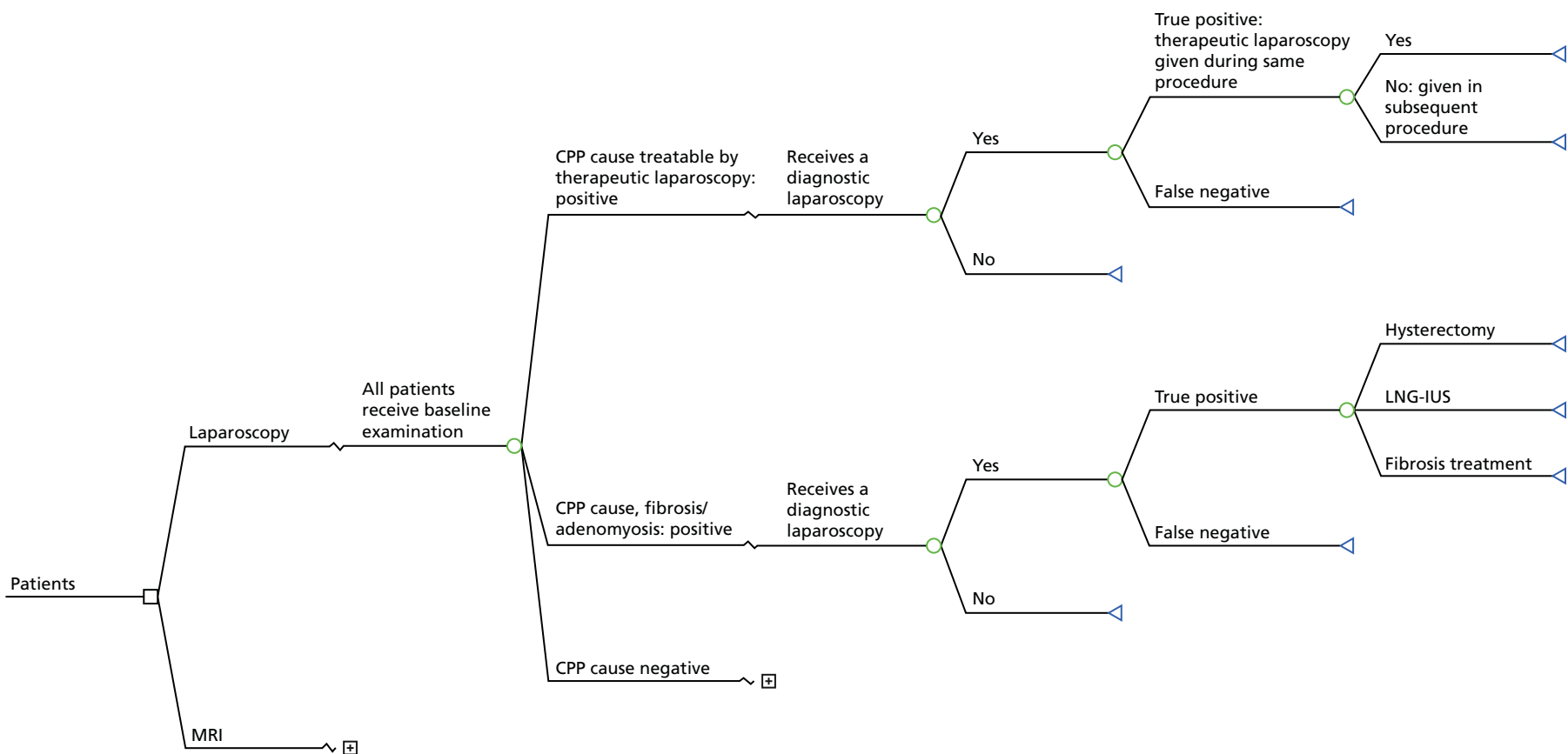


FIGURE 7 Model structure for the laparoscopy treatment arm for women who tested positive for a structural cause of CPP. LNG-IUS, levonorgestrel-releasing intrauterine system.

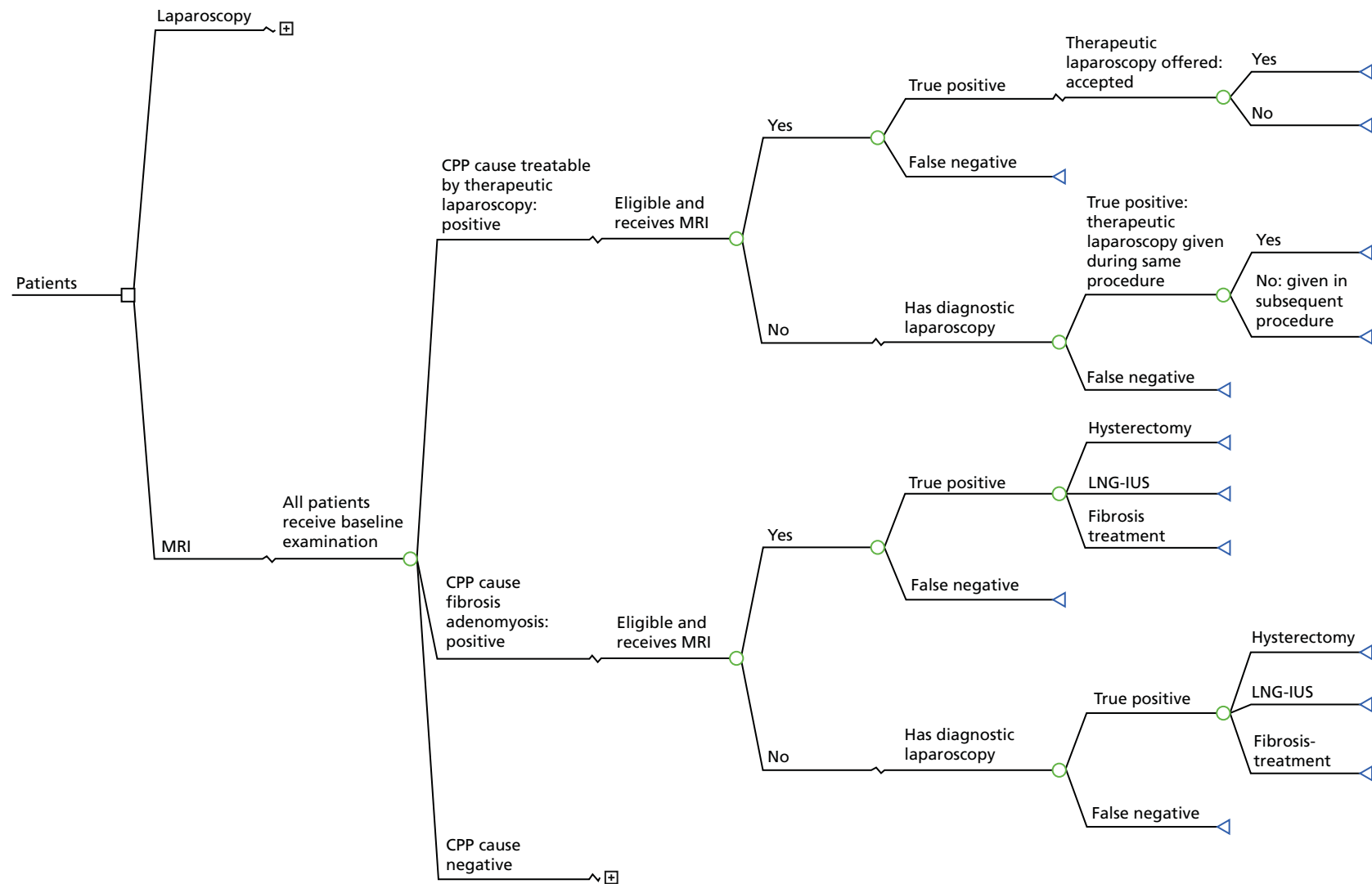


FIGURE 8 Model structure for the MRI treatment arm for women who tested positive for a structural cause of CPP. LNG-IUS, levonorgestrel-releasing intrauterine system.

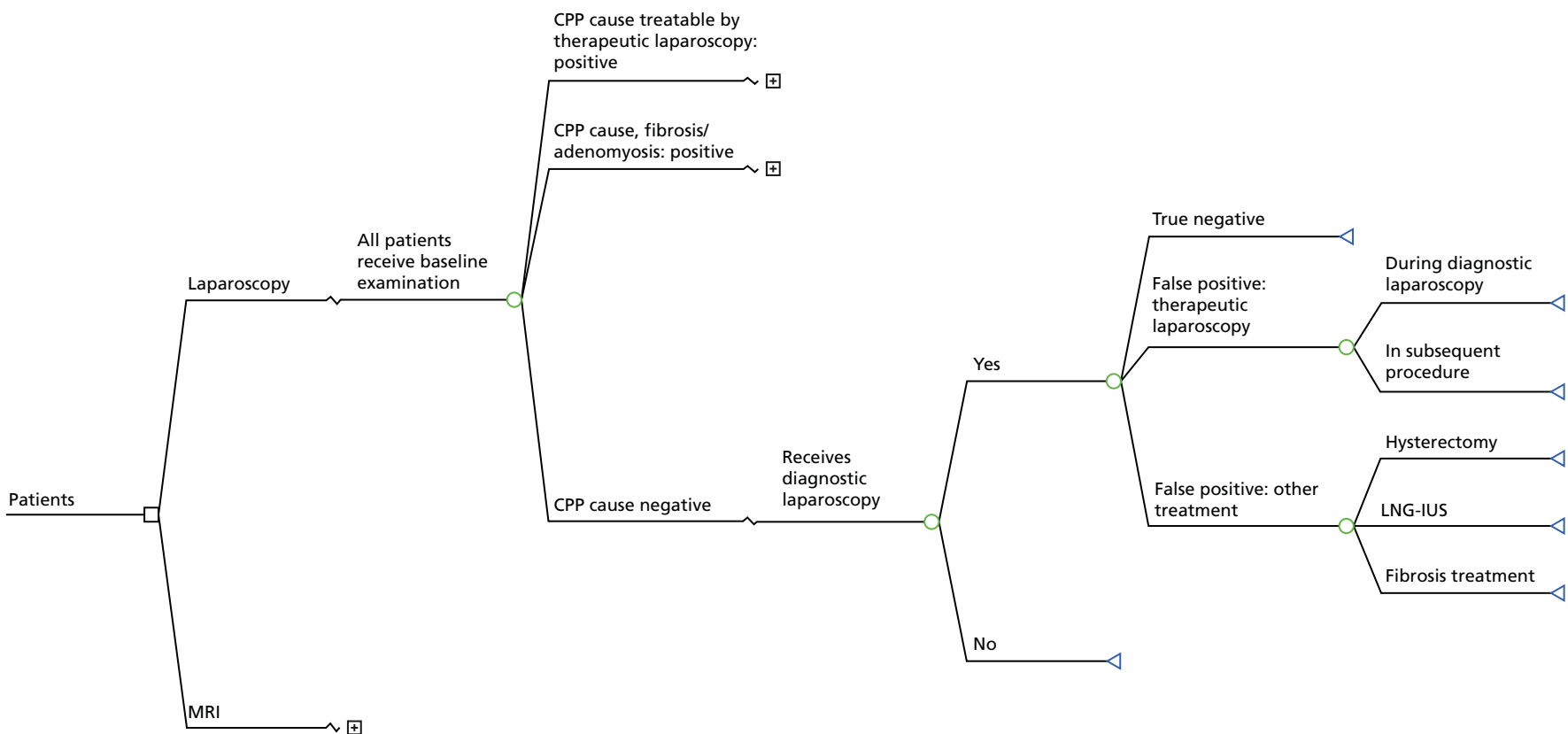


FIGURE 9 Model structure for the laparoscopy treatment arm for women who tested negative for a structural cause of CPP. LNG-IUS, levonorgestrel-releasing intrauterine system.

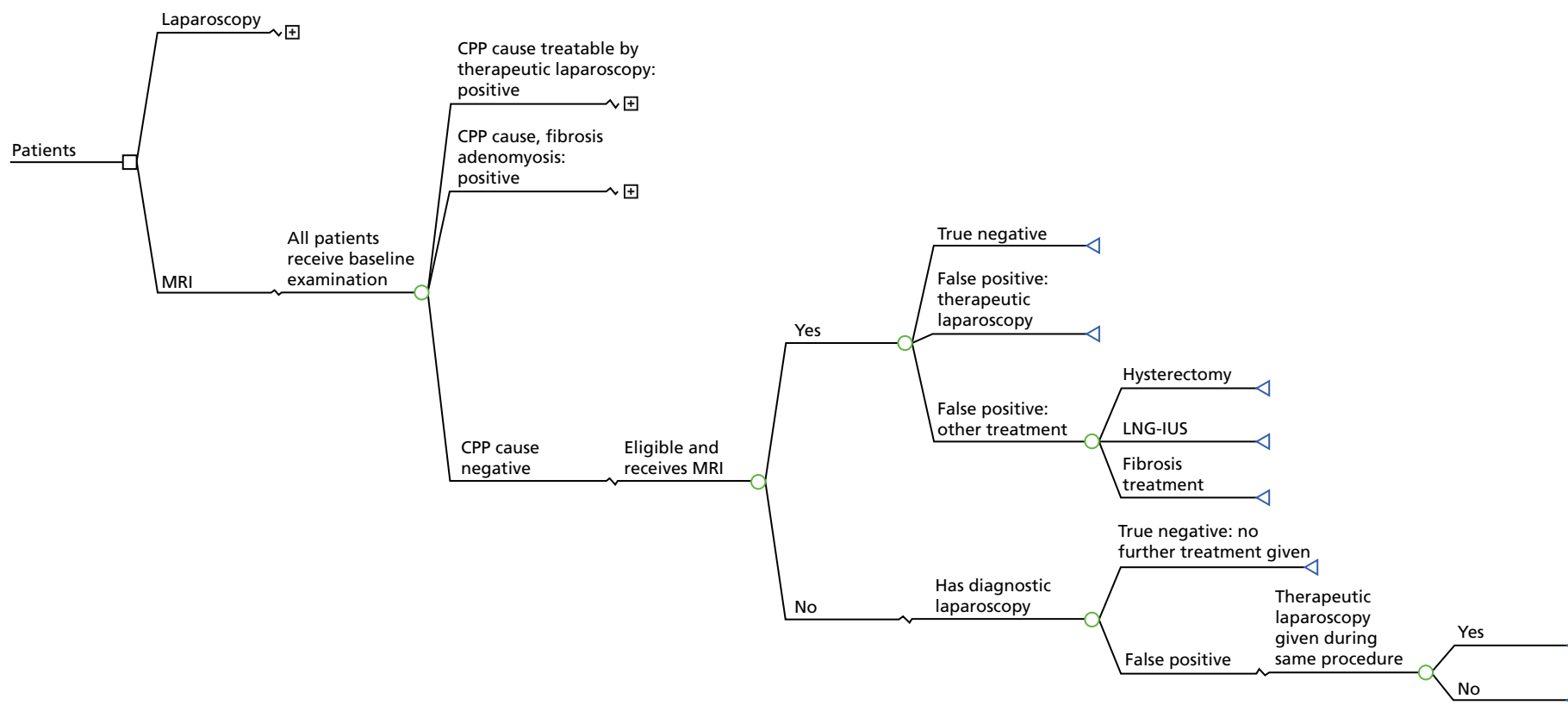


FIGURE 10 Model structure for the MRI treatment arm for women who tested negative for a structural cause of CPP. LNG-IUS, levonorgestrel-releasing intrauterine system.

when clinically indicated. The remainder of the pathway followed either the laparoscopy pathway or the MRI pathway.

Laparoscopy pathway

Women were offered a diagnostic laparoscopy, which may or may not have been accepted. For those women who received the diagnostic laparoscopy, this would give either a positive or a negative result for the presence of a structural cause of CPP. For those who tested positive for a cause that required a therapeutic laparoscopy, this would be given during the same or a subsequent procedure. For women who tested positive for a structural cause that required other treatment, in this case fibroids or adenomyosis, this treatment would be subsequently administered. For those women who tested negative, no further tests or treatment would be administered, and if this was a false-negative test result, then the patient would remain in a decreased state of health (see *Figures 7 and 9*).

Magnetic resonance imaging pathway

Women were offered a MRI scan that may or may not have been accepted; women who rejected or were ineligible for a MRI scan would then follow the laparoscopy pathway (see *Figures 8 and 9*). Women who received a MRI scan and a positive test result were given a therapeutic laparoscopy or other treatment, as appropriate. For women who tested negative, no further tests or treatments were administered and, again, if the test result was false negative, these women would remain in a decreased state of health (see *Figure 10*).

No testing (neither magnetic resonance imaging nor laparoscopy)

A no-testing scenario was incorporated into this analysis, which was assumed to consist of a baseline examination (see *Model assumptions*). Although we acknowledge that this strategy would not be considered in practice for this patient group, the inclusion of no testing is important for comparative purposes, and can show whether or not a testing strategy is more harmful than no testing.

Model assumptions

In order to implement a working model structure and enable this analysis to be carried out, a number of model assumptions were necessary. These assumptions were broadly categorised as being related to resource use/costs, diagnostic pathway, course of treatment and QoL, and are described, as follows, in the next sections.

Resource use/costs

- The baseline examination given to all women at the start of each strategy (including no testing) consisted of the following procedures:
 - patient history taken by a research nurse and consultant, which takes 20 minutes
 - physical examination undertaken by a consultant, which takes 20 minutes
 - an abdominal and vaginal ultrasound (for two-thirds of women).
- A patient would receive one consultation prior to each laparoscopy, and then one consultation post laparoscopy, each taking 20 minutes with a consultant.
- To allow for the increased likelihood of complications related to therapeutic laparoscopy, therapeutic laparoscopy was assumed to cost 10% more than diagnostic laparoscopy.

Diagnostic pathway

- Only women with a structural cause of CPP were regarded as having a true-positive diagnosis and being in need of treatment.
- A patient who received a false-positive result from a diagnostic laparoscopy would receive therapeutic laparoscopy during the same or a subsequent procedure.
- A patient who received a false-negative test result would receive no further testing or treatment and would remain in a decreased state of health for the remainder of the time horizon of the analysis.

- Women would always accept a laparoscopy following a positive MRI result.
- A patient would always accept a diagnostic laparoscopy after having refused or having been ineligible for a MRI scan.

Course of treatment

- For women with adenomyosis, 50% had a hysterectomy and 50% had a levonorgestrel-releasing intrauterine system.
- Among those with deep-infiltrating endometriosis, 50% would have a second therapeutic laparoscopy.
- It was assumed that all women only required one type of treatment, with women who had more than one condition being assumed to need a therapeutic laparoscopy if any one of their conditions required this treatment.

Quality of life

- Women who dropped out after the baseline assessment, and did not go on to receive testing or treatment, had the same QoL estimates at 6 months as were measured at baseline.
- For women who received treatment, their QoL changed in response to treatment 4 weeks after entering the model.
- The impact on QoL of therapeutic laparoscopy was assumed to be the same as for all other treatments.

Data requirements

The data requirements for the economic evaluation were fulfilled through the use of the patient-level data collected as part of the MEDAL study.

Prevalence of the structural causes of chronic pelvic pain

Only women who were found to have a structural cause of CPP were considered to require treatment, with the remaining women assumed to be negative for structural causes. In this model, the treatments for structural causes of CPP were presented in two groups:

1. therapeutic laparoscopy – for women who required a therapeutic laparoscopy to resolve their structural cause of CPP
2. other treatment – for women who had a structural cause of CPP, but did not require a therapeutic laparoscopy, which, in this case, are women with fibroids and adenomyosis.

As shown in *Table 14*, data collected during the MEDAL study found that some women had more than one structural cause. In this case, if any of the women's structural causes were considered to be an appropriate justification for a therapeutic laparoscopy, they would be incorporated into this group (see *Model assumptions*). The prevalence of structural causes by treatment group is shown in *Table 15*.

Test accuracy

The test accuracy of diagnostic laparoscopy and MRI was based on the sensitivity and specificity for the detection of structural causes that required either a therapeutic laparoscopy or other treatment, and were informed by data collected during the MEDAL study.

TABLE 14 Disease status of the women in the MEDAL study by treatment group

Condition	Number of cases	Prevalence, % (95% CI) (N = 287)	Treatment group
None	164	57.1 (51.2 to 62.9)	–
Superficial peritoneal endometriosis only	46	16.0 (12.0 to 20.8)	TL
Deep-infiltrating endometriosis only	20	7.0 (4.3 to 10.6)	TL
Adhesions only	15	5.2 (3.0 to 8.5)	TL
Superficial peritoneal endometriosis and deep-infiltrating endometriosis	10	3.5 (1.7 to 6.3)	TL
Adenomyosis only	8	2.8 (1.2 to 5.4)	Other
Superficial peritoneal endometriosis and adhesions	5	1.7 (0.6 to 4.0)	TL
Superficial peritoneal endometriosis, deep-infiltrating endometriosis and endometrioma of the ovary	4	1.4 (0.4 to 3.5)	TL
Fibroids only	3	1.1 (0.2 to 3.0)	Other
Deep-infiltrating endometriosis and endometrioma of the ovary	2	0.7 (0.1 to 2.5)	TL
Endometrioma of the ovary only	2	0.7 (0.1 to 2.5)	TL
Superficial peritoneal endometriosis and endometrioma of the ovary	2	0.7 (0.1 to 2.5)	TL
Deep-infiltrating endometriosis and adhesions	1	0.4 (0.01 to 1.9)	TL
Deep-infiltrating endometriosis, superficial peritoneal endometriosis and adhesions	1	0.4 (0.01 to 1.9)	TL
Deep-infiltrating endometriosis, superficial peritoneal endometriosis, adhesions and fibroids	1	0.4 (0.01 to 1.9)	TL
Deep-infiltrating endometriosis, superficial peritoneal endometriosis, endometrioma of the ovary, adhesions and adenomyosis	1	0.4 (0.01 to 1.9)	TL
Ovarian cysts and adenomyosis	1	0.4 (0.01 to 1.9)	TL
Superficial peritoneal endometriosis and adenomyosis	1	0.4 (0.01 to 1.9)	TL

Other, other treatment; TL, therapeutic laparoscopy.

TABLE 15 Prevalence of structural causes stratified by treatment group

Prevalence category	Number of cases (% of total)	Notes
No structural cause	164 (57.14)	
Structural cause requires therapeutic laparoscopy ^a	112 (39.02)	
Structural cause requires other treatment ^b	11 (3.83)	
Total	287	

a Includes women with superficial peritoneal endometriosis, deep-infiltrating endometriosis (who may require a second therapeutic laparoscopy), endometrioma of the ovary, adhesions and ovarian cysts.

b Includes women with fibroids or adenomyosis (no women with both conditions were seen in this study). Percentages do not add up to 100% as a result of rounding.

The test accuracy of MRI and diagnostic laparoscopy, respectively, was explored through the use of ROC curves for structural causes that required a therapeutic laparoscopy, and for those that required other treatment (*Figure 11*). The ROC curve was created by plotting the true-positive rate (true-positive rate or sensitivity) against the false-positive rate (false-positive rate or $1 - \text{specificity}$) at various threshold settings.

In our analysis, we refer to a low cut-off point as one that sets a high sensitivity and low specificity (more positive tests), and a high cut-off point as one that sets a low sensitivity and high specificity (more negative tests).

Calculation of test-accuracy parameters

In order to incorporate the impact of the test results on patient outcomes into the model, it was necessary to combine the prevalence of CPP with the sensitivity and specificity of the tests.

A patient in the model could be in one of three disease states:

1. patient has a structural cause that requires therapeutic laparoscopy
2. patient has a structural cause that requires other treatment
3. patient tests negative for any structural cause of CPP.

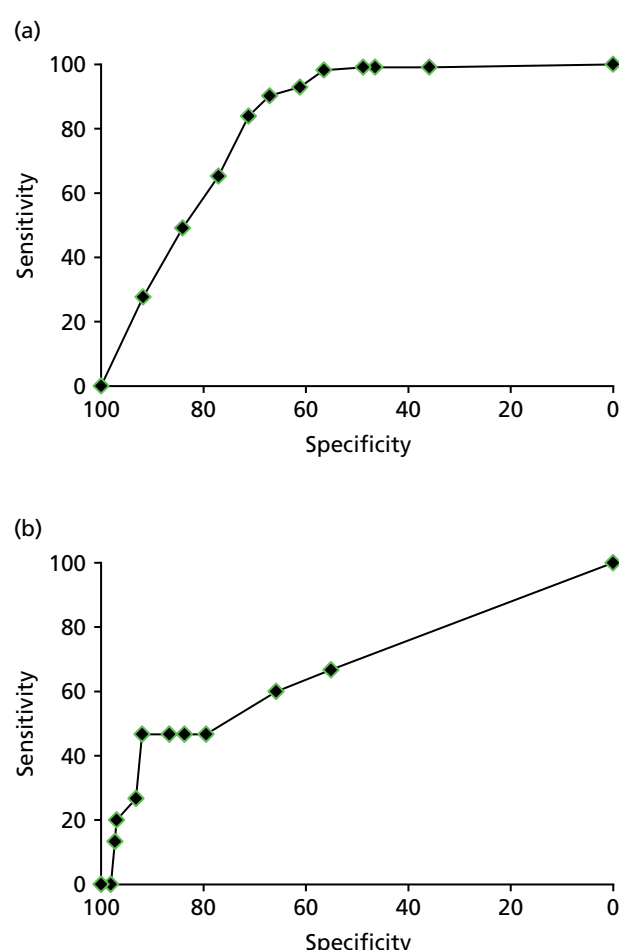


FIGURE 11 Receiver operating characteristic curves for laparoscopic and MRI diagnoses. (a) Diagnostic laparoscopy for structural causes of CPP that required treatment using a therapeutic laparoscopy; (b) diagnostic laparoscopy for structural causes of CPP that required other treatment (not therapeutic laparoscopy); (c) MRI scan for structural causes of CPP that required treatment using a therapeutic laparoscopy; and (d) MRI scan for structural causes of CPP that required other treatment (not therapeutic laparoscopy). (*continued*)

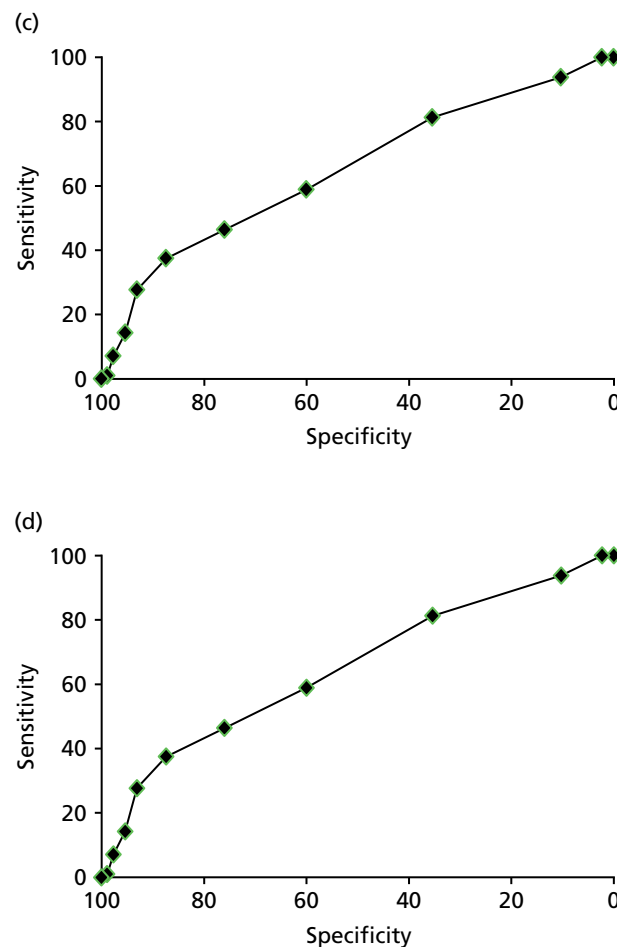


FIGURE 11 Receiver operating characteristic curves for laparoscopic and MRI diagnoses. (a) Diagnostic laparoscopy for structural causes of CPP that required treatment using a therapeutic laparoscopy; (b) diagnostic laparoscopy for structural causes of CPP that required other treatment (not therapeutic laparoscopy); (c) MRI scan for structural causes of CPP that required treatment using a therapeutic laparoscopy; and (d) MRI scan for structural causes of CPP that required other treatment (not therapeutic laparoscopy).

The prevalence of disease based on these groupings was combined with the test-accuracy characteristics of the tests through the following equations to give the proportion of women who test true or false positive or negative for a structural cause of CPP:

$$TP(TL) = \text{Sens}(TL)P(TL), \quad (1)$$

$$FN(TL) = [1 - \text{Sens}(TL)]P(TL), \quad (2)$$

$$TP(OT) = \text{Sens}(OT)P(OT), \quad (3)$$

$$FN(OT) = [1 - \text{Sens}(OT)]P(OT), \quad (4)$$

$$TN = [1 - P(TL) - P(OT)]\text{Spec}(TL)\text{Spec}(OT), \quad (5)$$

$$FP(TL) = (((1 - P(TL) - P(OT)) - TN)(Spec(OT)/(Spec(OT) + Spec(TL))), \quad (6)$$

$$FP(OT) = (((1 - P(TL) - P(OT)) - TN)Spec(TL)/(Spec(OT) + Spec(TL))), \quad (7)$$

in which Sens is the sensitivity of the test, Spec is the specificity of the test, TP is a true-positive test result, FP is false positive, TN is true negative, FN is false negative and P is the prevalence of women with a structural cause subdivided into those women who require a therapeutic laparoscopy (TL) and those who require other treatment (OT).

Further model parameters

Parameters describing the uptake of MRI and laparoscopy are shown in *Table 16* below. These were informed by the data collection conducted in the MEDAL study. As described in *Model assumptions*, 50% of women with deep-infiltrating endometriosis were assumed to receive two therapeutic laparoscopies. This was incorporated into the analysis for the purpose of calculating costs, but had no influence on patient outcomes, and the impact of this assumption was examined during the sensitivity analysis (see *Sensitivity analyses*).

Cost and resource data

All costs used in this evaluation are in UK pounds sterling (at the 2013 value). *NHS Reference Costs 2013–14*⁸⁷ were used to attribute costs to the resource use. The *Unit Costs of Health and Social Care 2014*⁸⁸ from the Personal Social Services Research Unit were used to obtain the costs for staff time. The resource costs are shown in *Table 17*.

Outcomes

The primary outcome of this economic evaluation was the quality-adjusted life-year (QALY). In addition, the outcomes of cost per patient correctly diagnosed, and cost per patient correctly treated, were examined.

The QALY outcome measure was informed by QoL estimates obtained from the MEDAL study, through the EQ-5D-3L questionnaire, which was administered at baseline and 6 months post laparoscopy. This was not a randomised controlled trial in which women followed different patient pathways, and so it was not possible to obtain the QoL values for all the different patient pathways. Instead, the baseline values were used to inform the impact of a patient either having a structural cause of CPP or not having one, and then 6-month values were used to inform the impact on QoL of having a treatment with or without a cause of CPP. As no information was collected describing the impact of treatment on fibrosis or adenomyosis at 6 months, only the impact of therapeutic laparoscopy on QoL was considered. This QoL effect was assumed to be the same as the impact of treatment for fibrosis or adenomyosis (see *Model assumptions*). The QoL values used in this study are shown in *Table 18*.

TABLE 16 Parameters used in the decision model

Parameters	Value	References	Notes
Proportion of eligible women who received a laparoscopy when offered	288/309	Data	Intraoperative form missing for one patient, and so was not included in the final analysis
Proportion of eligible women who received a MRI scan when offered	291/329	Data	Intraoperative form missing for four women, and so were not included in the final analysis
Proportion of women who had a therapeutic laparoscopy during the same procedure as a diagnostic laparoscopy	92/112	Data	Remainder of women had therapeutic laparoscopy during a subsequent procedure

TABLE 17 Breakdown of the costs used in the economic analysis

Item	Code	Cost (£)	Reference	Notes
Consultant/hour	Consultant medical (with qualification costs)	140	<i>Unit Costs of Health and Social Care 2014</i> ⁸⁸	
Nurse/hour	Nurse, day ward [includes staff nurse, registered nurse, registered practitioner]/ hour of patient contact (with qualification costs)]	100		
Ultrasound	RA24Z ultrasound scan, ≥ 20 minutes (diagnostic imaging)	59	<i>NHS Reference Costs 2013–14</i> ⁸⁷	
Baseline examination	Twenty minutes of consultation time with a consultant and a nurse, followed by 20 minutes of physical examination by a consultant, and then two-thirds of women received an ultrasound scan	166		Assume mean is equal to standard error for nurse and consultant time ⁸⁹
Diagnostic laparoscopy	MA10Z minor laparoscopic or endoscopic procedure, upper genital tract procedure (day case)	1217	<i>NHS Reference Costs 2013–14</i> ⁸⁷	Therapeutic laparoscopy assumed to be 10% more expensive than diagnostic laparoscopy (see assumptions)
MRI	RA03Z MRI scan, one area, pre and post contrast (diagnostic imaging)	203		
Hysterectomy	MA07G major open upper genital tract procedures with a CC score of 0–2	3299		Applied to 50% of those women with adenomyosis
Implantation of an intrauterine device	MA35Z implementation of an intrauterine device	311		Applied to 50% of those women with adenomyosis
Fibroids	MB09C non-malignant gynaecological disorders with interventions with a CC score of 0–2	2520		

CC, complications and comorbidities.
All costs are in pounds sterling (£); the cost year is 2013/14.

Analysis

This model-based economic evaluation takes the form of a cost–utility analysis based on the primary outcome of cost per QALY. The secondary outcomes of the cost per patient correctly diagnosed and cost per patient appropriately treated were also considered as part of the wider cost-effectiveness analyses. The perspective of the analysis is the health-care provider perspective (the UK NHS) in a secondary care (hospital) setting. Given the 6-month time horizon, no discounting was applied. The results in this study are described using the incremental cost-effectiveness ratio (ICER); this is defined as the difference between the costs of the two strategies divided by the difference in their outcomes.

In this analysis, for each of the strategies, MRI and laparoscopy, the cut-off value for sensitivity and specificity is treated as a decision variable, rather than an endogenous characteristic of each test. Given that there are 12 cut-off points for each ROC curve, this leads to 24 alternative strategies for MRI and laparoscopy, and, with the no-testing strategy, 25 alternative strategies were considered here.

TABLE 18 Quality-of-life values used to inform the QALY outcome

QoL estimate	Time point, EQ-5D-3L mean score (SD, n)		Application in the model
	Baseline	6 months	
Have a cause of CPP that requires therapeutic laparoscopy ^a	0.56 (0.32, 112)		QoL of women with a cause of CPP at baseline, and those who tested FN and, thus, were not treated
Do not have a cause of CPP ^b	0.53 (0.32, 151)		QoL of women who tested negative for a cause of CPP at baseline
Do not have any cause of CPP and receive therapeutic laparoscopy ^c		0.55 (0.35, 53)	QoL of negative women who tested FP and received unnecessary treatment
Do not have any cause of CPP and do not receive therapeutic laparoscopy ^d		0.69 (0.26, 71)	QoL of negative women at 6 months who tested TN and did not receive treatment
Have cause of CPP that requires therapeutic laparoscopy and receives therapeutic laparoscopy ^e		0.74 (0.28, 68)	QoL of women who tested TP and received treatment

FN, false negative; FP, false positive; TN, true negative.
a Twenty-eight values missing from 6 months post laparoscopy.
b Five values missing from baseline.
c Eight values missing from 6 months post laparoscopy.
d Twenty-three values missing from 6 months post laparoscopy.
e Twenty-four values missing from 6 months post laparoscopy.

Sensitivity analyses

A range of one-way sensitivity analyses were undertaken to gain further insights into the impact of reasonable changes to key parameters in the model, as follows:

1. To examine the impact of the 6-month time horizon on the model results, this was extended beyond the period of data collection up to 3 years. However, as a result of a lack of data being collected in this study beyond 6 months, this required two assumptions to be made. The first was that the QoL informed by data at 6 months was constant up to the time horizon under consideration, and the second was that no costs were incurred beyond the 6 months.
2. To examine the impact of changing the prevalence of structural causes among women with CPP in this model, this was varied between 0% and 100%. However, when applying the sensitivity analysis, it was assumed that the ratio of positive women in the therapeutic laparoscopy group to the other treatment group remained the same, as informed by the MEDAL study data.
3. To examine the impact of the assumption that 50% of women with deep-infiltrating endometriosis required two therapeutic laparoscopies, this value was varied between 0% and 100%.
4. To examine the impact of the assumption that 50% of women with adenomyosis had a hysterectomy and 50% had a levonorgestrel-releasing intrauterine system, these values were varied from 0% to 100% (always adding up to 100%).

In order to incorporate the uncertainty that is inherent in many of the parameters in this analysis, a probabilistic sensitivity analysis (PSA) was undertaken, with each parameter value being assigned a distribution and the model being run 10,000 times, sampling on each occasion from these distributions. The distributions applied to the parameters for the PSA are shown in *Appendix 3*.

Results

This section starts by examining the cost-effectiveness of MRI, laparoscopy, and no testing, whereby the different cut-off values to inform the sensitivity and specificity for each test was treated as a separate strategy. At baseline, a 6-month time horizon was utilised.

Quality-adjusted life-year outcome

Using the outcome of the QALYs, the cost-effectiveness of MRI and laparoscopy across varying cut-off points is shown in *Table 19*.

As shown in *Table 20*, the baseline results for strategies on the cost-effectiveness frontier, using the outcome of the QALY for most scenarios and excluding the scenarios for the more extreme cut-off

TABLE 19 Baseline deterministic cost-effectiveness results for MRI and laparoscopy at various cut-off points using the QALY outcome and a 6-month time horizon

Strategy and cut-off point ^a	Cost (£)	QALY gain
MRI – 2	1856	0.3077
MRI – 1	1908	0.3082
MRI – 0	2097	0.3083
Laparoscopy – 0	2231	0.3084
MRI – 10	538	0.3094
No testing	166	0.3095
MRI – 11	497	0.3095
Laparoscopy – 11	1387	0.3095
MRI – 3	1691	0.3097
MRI – 4	1397	0.3100
MRI – 9	623	0.3103
MRI – 8	729	0.3109
MRI – 5	1171	0.3121
Laparoscopy – 1	2017	0.3134
MRI – 7	862	0.3137
MRI – 6	1002	0.3140
Laparoscopy – 10	1481	0.3140
Laparoscopy – 2	1973	0.3166
Laparoscopy – 9	1593	0.3176
Laparoscopy – 3	1904	0.3188
Laparoscopy – 8	1664	0.3200
Laparoscopy – 4	1873	0.3212
Laparoscopy – 5	1839	0.3215
Laparoscopy – 7	1747	0.3227
Laparoscopy – 6	1793	0.3235

^a Cut-off points are based on the data points in *Figure 11*, which were obtained from questionnaires. There are 12 data points in these graphs, which matches the number of cut-off points.

TABLE 20 Baseline deterministic results for strategies on the cost-effectiveness frontier using the QALY outcome

Strategy	Cost (£)	Incremental cost (£)	QALYs gained	Incremental effectiveness	ICER (cost/QALY), £
No testing	166		0.3095		
Laparoscopy (cut-off value of 6)	1793	1627	0.3235	0.014	116,618

points, shows that laparoscopy is more effective than MRI in terms of QALYs gained. For both MRI and laparoscopy, for some of the highest and lowest cut-off values that appear at the extreme ends on the ROC curve, in terms of QALYs gained, these are no better than no testing, and, in some cases, are even worse than no testing. This can be explained, as, for some of these strategies, women are given expensive unnecessary treatment as a result of false-negative test results. Finally, it is noted that for most scenarios, excluding the scenarios for the more extreme cut-off points, laparoscopy is more effective than MRI in terms of QALYs gained. The cost-effectiveness frontier for the results in *Table 20* is shown in *Figure 12*.

As can be seen in *Figure 12*, all the strategies were either dominated (less effective and more costly) or subject to extended dominance (another strategy provides more units of benefit at a lower cost per unit of benefit), except for no testing and laparoscopy, at a cut-off value of 6, which therefore lies on the cost-effectiveness frontier.

As shown in *Table 20*, for a 6-month time horizon, using the outcome of the QALY, laparoscopy (with a cut-off value of 6) is more expensive (£1793 vs. £166) and more effective (0.3235 vs. 0.3095) than no testing. In addition, the ICER for laparoscopy (with a cut-off value of 6) compared with no testing is £116,618, which means that one extra QALY gained from laparoscopy (with a cut-off value of 6) will cost £116,618 compared with no testing.

Correctly diagnosed outcome

Using the outcome measure of a correct diagnosis, the cost-effectiveness results for MRI and laparoscopy at various cut-off points, for a 6-month time horizon, are shown in *Table 21*.

As shown in *Table 21*, for the outcome of correctly diagnosed, the strategies at the extreme ends of the ROC curve were the least effective, with a few exceptions. The most effective strategies were the laparoscopy strategies around the mid-point of the ROC curve, that is, the cut-off values 4–9. The cost-effectiveness frontier for the results in *Table 21* is shown in *Figure 13*.

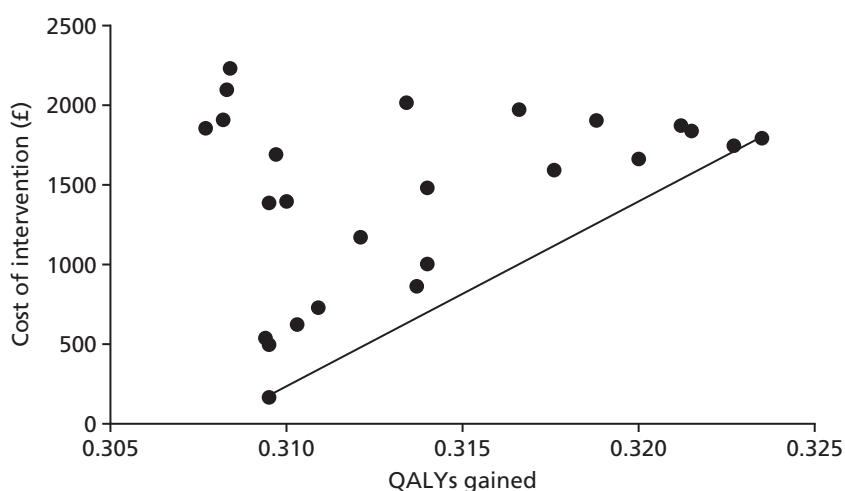
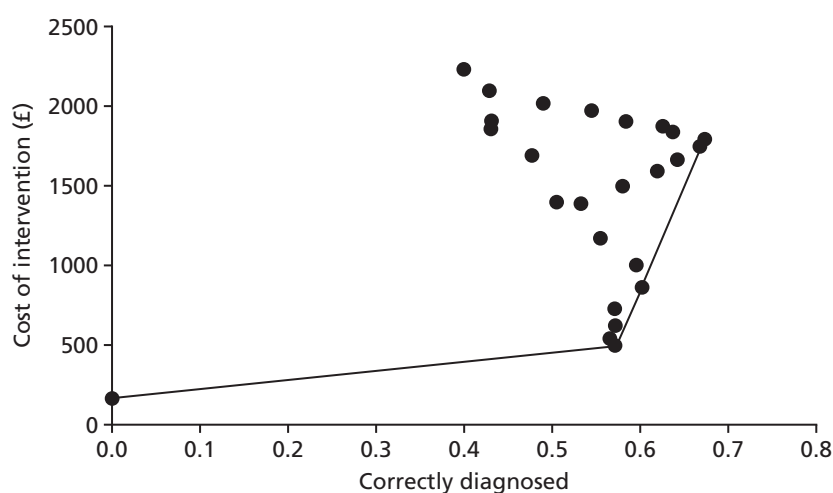
**FIGURE 12** Cost-effectiveness frontier for the baseline strategies, using the QALY outcome and the 6-month time horizon.

TABLE 21 Baseline deterministic cost-effectiveness results for MRI and laparoscopy at various cut-off points, using the correctly diagnosed outcome and a 6-month time horizon

Strategy and cut-off point ^a	Cost (£)	Correctly diagnosed
No testing	166	0.0000
Laparoscopy – 0	2231	0.3994
MRI – 0	2097	0.4286
MRI – 2	1856	0.4303
MRI – 1	1908	0.4310
MRI – 3	1691	0.4770
Laparoscopy – 1	2017	0.4897
MRI – 4	1397	0.5051
Laparoscopy – 11	1387	0.5326
Laparoscopy – 2	1973	0.5447
MRI – 5	1171	0.5547
MRI – 10	541	0.5654
MRI – 8	729	0.5709
MRI – 11	497	0.5714
MRI – 9	623	0.5717
Laparoscopy – 10	1498	0.5801
Laparoscopy – 3	1904	0.5838
MRI – 6	1002	0.5956
MRI – 7	862	0.6022
Laparoscopy – 9	1593	0.6195
Laparoscopy – 4	1873	0.6255
Laparoscopy – 5	1839	0.6369
Laparoscopy – 8	1664	0.6421
Laparoscopy – 7	1747	0.6679
Laparoscopy – 6	1793	0.6733

^a Cut-off points are based on the data points in *Figure 11*, which were obtained from questionnaires. There are 12 data points in these graphs, which matches the number of cut-off points.

**FIGURE 13** Cost-effectiveness frontier for baseline strategies, using the correctly diagnosed outcome and a 6-month time horizon.

As can be seen in *Figure 13*, all the strategies were either dominated or subject to extended dominance, except for the four strategies on the cost-effectiveness frontier, these being no testing, MRI (cut-off value of 11), MRI (cut-off value of 6) and laparoscopy (cut-off value of 6).

As shown in *Table 22*, for a 6-month time horizon and using the outcome of correctly diagnosed, MRI (cut-off value of 11) was more expensive and more effective than no testing. MRI (cut-off value of 7) was more expensive and more effective than MRI (cut-off value of 11), and laparoscopy (cut-off value of 6) was more expensive and more effective than MRI (cut-off value of 7). In the case of the comparison between laparoscopy (cut-off value of 6) and MRI (cut-off value of 7), the ICER was £13,092, which means that it would cost £13,092 for one extra correct diagnosis for laparoscopy (cut-off value of 6) compared with MRI (cut-off value of 7).

Correctly treated outcome

Using the outcome measure of a patient being positive for a structural cause of CPP and receiving the correct treatment, the model results for MRI and laparoscopy at various cut-off points, for a 6-month time horizon, are shown in *Table 23*.

Table 23 shows that the least effective scenarios were the MRI strategies with low cut-off points, that is, those with high sensitivity and low specificity. The most effective scenarios were the laparoscopy scenarios for the cut-off values 3–10. The cost-effectiveness frontier for the results in *Table 23* is shown in *Figure 14*.

As can be seen in *Figure 14*, all the strategies were either dominated or subject to extended dominance, except for the three strategies on the cost-effectiveness frontier, which were no testing, MRI (cut-off value of 11), and laparoscopy (cut-off value of 6).

As shown in *Table 24*, for a 6-month time horizon, using the outcome of a women being correctly treated with MRI (cut-off value of 11) was more expensive (£497 vs. £166) and more effective (0.5714 vs. 0) than no testing, and the ICER for MRI (cut-off value of 11) compared with no testing was £579, which means that one extra patient correctly treated with MRI (cut-off value of 11) would cost £579 compared with no testing. Likewise, laparoscopy (cut-off value of 6) was more expensive (£1793 vs £497) and more effective (0.7121 vs. 0.5714) than MRI (cut-off value of 11) with an ICER of £9211, meaning that one extra woman correctly treated with laparoscopy (cut-off value of 6) will cost £9211 compared with MRI (cut-off value of 11).

TABLE 22 Baseline deterministic results for strategies on the cost-effectiveness frontier, using the correctly diagnosed outcome

Strategy	Cost (£)	Incremental cost (£)	Correctly diagnosed	Incremental effectiveness	ICER (cost/correctly diagnosed), £
No testing	166		0		
MRI (cut-off value of 11)	497	331	0.5714	0.5714	579
MRI (cut-off value of 7)	862	365	0.6022	0.0308	11,864
Laparoscopy (cut-off value of 6)	1793	1627	0.6733	0.0711	13,092

TABLE 23 Baseline deterministic cost-effectiveness results for MRI and laparoscopy at various cut-off points, using the correctly treated outcome and a 6-month time horizon

Strategy and cut-off value ^a	Cost (£)	Correctly treated
No testing	166	0.0000
MRI – 0	2097	0.4286
MRI – 2	1856	0.4303
MRI – 1	1908	0.4310
Laparoscopy – 0	2231	0.4383
MRI – 3	1691	0.4770
MRI – 4	1397	0.5051
Laparoscopy – 1	2017	0.5285
MRI – 5	1171	0.5547
MRI – 10	541	0.5654
MRI – 8	729	0.5709
MRI – 11	497	0.5714
Laparoscopy – 11	1387	0.5714
MRI – 9	623	0.5717
Laparoscopy – 2	1973	0.5835
MRI – 6	1002	0.5956
MRI – 7	862	0.6022
Laparoscopy – 10	1498	0.6189
Laparoscopy – 3	1904	0.6226
Laparoscopy – 9	1593	0.6583
Laparoscopy – 4	1873	0.6643
Laparoscopy – 5	1839	0.6757
Laparoscopy – 8	1664	0.6810
Laparoscopy – 7	1747	0.7068
Laparoscopy – 6	1793	0.7121

^a Cut-off points are based on the data points in *Figure 11*, which were obtained from questionnaires. There are 12 data points in these graphs, which matches the number of cut-off points.

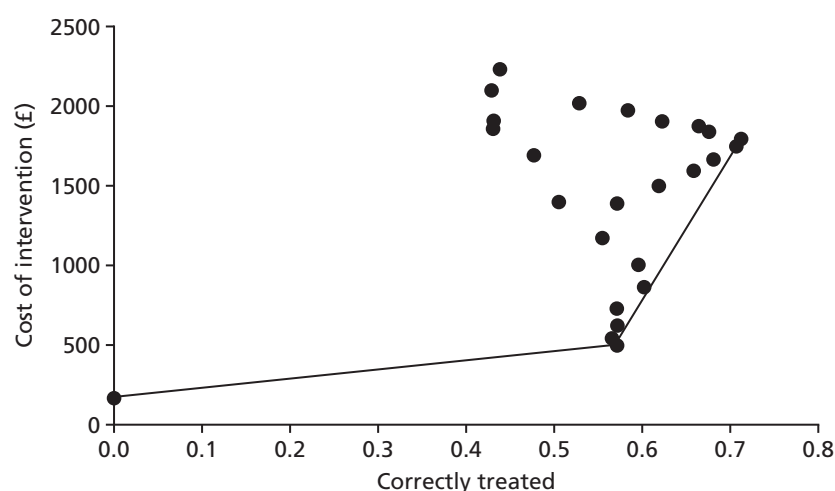
**FIGURE 14** Cost-effectiveness frontier for baseline strategies, using the correctly treated outcome with a 6-month time horizon.

TABLE 24 Baseline deterministic results for strategies on the cost-effectiveness frontier, using the correctly treated strategy outcome

Strategy	Cost (£)	Incremental cost (£)	Correctly diagnosed	Incremental effectiveness	ICER (cost/correctly treated), £
No testing	166		0		
MRI (cut-off value of 11)	497	331	0.5714	0.5714	579
Laparoscopy (cut-off value of 6)	1793	1296	0.7121	0.1407	9211

Sensitivity analysis

Time horizon

The baseline analysis has a time horizon of 6 months. The impact of extending the time horizon up to 3 years for the outcome of the QALY on model results is shown in *Table 25*.

The results when varying the time horizon are shown in *Table 25*. In each case, only the strategies that lie on the cost-effectiveness plane are shown (dominated and subject to extended dominance scenarios not shown). It can be seen that, as the time horizon of the analysis is increased, laparoscopy (cut-off value of 6) becomes more cost-effective than no testing. However, it is noted that, although the effectiveness of the interventions increases over time, as a result of the model assumptions, the costs remain the same.

Prevalence

The impact of varying the overall prevalence of structural causes among women with CPP was examined, with the results shown in *Table 26*. Note that only the results for the scenarios that lie on the cost-effectiveness frontier are described.

TABLE 25 Model results for the QALY outcome with variations in the time horizon of the analysis strategy

Time horizon and strategy	Cost (£)	Incremental cost (£)	QALYs gained	Incremental effectiveness	ICER (cost/QALY), £
6-month time horizon					
No testing	166		0.3095		
Laparoscopy (cut-off value of 6)	1793	1627	0.3235	0.0140	116,618
1-year time horizon					
No testing	166		0.6267		
Laparoscopy (cut-off value of 6)	1793	1627	0.6574	0.0307	53,000
2-year time horizon					
No testing	166		1.2610		
Laparoscopy (cut-off value of 6)	1793	1627	1.3251	0.0641	25,384
3-year time horizon					
No testing	166		1.8952		
Laparoscopy (cut-off value of 6)	1793	1627	1.9929	0.0977	16,654

TABLE 26 Model results using variations in the prevalence of a structural cause of CPP

Prevalence	Strategy (with cut-off value)	Cost (£)	Incremental cost (£)	QALY gain	Incremental QALY	ICER (cost/QALY), £
0%	No testing	166		0.3317		Dominates
10%	No testing	166		0.3264		Dominates
20%	No testing	166		0.3214		Dominates
30%	No testing	166		0.3162		
30%	Laparoscopy – 7	1730	1564	0.3198	0.0036	430,292
40%	No testing	166		0.3110		
40%	Laparoscopy – 6	1788	1622	0.3227	0.0117	138,407
Baseline (42.86%)	No testing	166		0.3095		
Baseline (42.86%)	Laparoscopy – 6	1793	1627	0.3235	0.0140	116,618
50%	No testing	166		0.3058		
50%	Laparoscopy – 6	1805	1639	0.3256	0.0197	83,032
60%	No testing	166		0.3007		
60%	Laparoscopy – 6	1820	1654	0.3285	0.0278	59,389
70%	No testing	166		0.2955		
70%	Laparoscopy – 6	1836	1670	0.3315	0.0360	46,398
70%	Laparoscopy – 4	1896	60	0.3327	0.0012	48,842
80%	No testing	166		0.2903		
80%	Laparoscopy – 4	1905	1739	0.3369	0.0467	37,255
80%	MRI – 0	2061	156	0.3388	0.0018	85,124
90%	No testing	166		0.2852		
90%	MRI – 1	1959	1793	0.3445	0.0593	30,248
90%	MRI – 0	2051	91	0.3467	0.0022	40,899
100%	No testing	166		0.2800		
100%	MRI – 1	1970	1804	0.3523	0.0723	24,966
100%	MRI – 0	2042	72	0.3550	0.0027	26,169

Table 26 shows the model results with variations in the prevalence of a structural cause of CPP. At low prevalence rates (0–20%), no testing is cheaper and more effective than all other strategies. As the prevalence increases, laparoscopy (cut-off value of 6) becomes increasingly cost-effective up to 60%, and then laparoscopy (cut-off value of 4) is preferred for prevalences of 70–80%. At the highest prevalence rates, MRI with a low cut-off value becomes the preferred option. This can be explained through the fact that, with a high prevalence rate, it is better to offer a cheap test that has a high sensitivity and low specificity, as there will be few false-positive test results.

Deep-infiltrating endometriosis

It was assumed that at baseline 50% of women with deep-infiltrating endometriosis required a second therapeutic laparoscopy. The results in Table 27 show the impact of varying this assumption from 0% to 100%.

TABLE 27 Model results with a variation in the proportion of women with deep-infiltrating endometriosis who required a second therapeutic laparoscopy

Proportion of women with deep-infiltrating endometriosis who required two therapeutic laparoscopies	Cost (£)	Incremental cost (£)	QALY gain	Incremental QALY	ICER (cost/QALY), £
0%					
No testing	166		0.3095		
Laparoscopy – 6	1730	1564	0.3235	0.014	112,109
Baseline					
No testing	166		0.3095		
Laparoscopy – 6	1793	1627	0.3235	0.014	116,618
100%					
No testing	166		0.3095		
Laparoscopy – 6	1856	1690	0.3235	0.014	121,126

As shown in *Table 26*, varying the parameter describing the percentage of women with deep-infiltrating endometriosis who required two therapeutic laparoscopies had very little impact on the ICER values, and did not impact on the conclusions drawn from the model. All the ICER values lie close to the baseline ICER of £116,618 and remain above the £20,000–30,000 acceptance threshold for the cost/QALYs, as recommended by the National Institute for Health and Care Excellence (NICE).⁹⁰

Hysterectomy for adenomyosis

It was assumed at baseline that 50% of women with adenomyosis are treated with a hysterectomy. The results in *Table 28* show the impact of varying this assumption from 0% to 100%.

As shown in *Table 28*, varying the parameter describing the percentage of women with adenomyosis who are treated with hysterectomy had very little impact on the ICER values, and did not impact on the conclusions drawn from the model. All the ICER values lie close to the baseline ICER of £116,618, and remain above the £20,000–30,000 acceptance threshold for the cost/QALYs, as recommended by NICE.

TABLE 28 Model results with variation in the proportion of women with adenomyosis who were treated with a hysterectomy

Proportion of women with adenomyosis who were treated with a hysterectomy	Cost (£)	Incremental cost (£)	QALY gain	Incremental QALY	ICER (cost/QALY), £
0%					
No testing	166		0.3095		
Laparoscopy – 6	1682	1516	0.3235	0.014	108,624
Baseline					
No testing	166		0.3095		
Laparoscopy – 6	1793	1627	0.3235	0.014	116,618
100%					
No testing	166		0.3095		
Laparoscopy – 6	1905	1739	0.3235	0.014	124,612

Probabilistic sensitivity analysis

The results for the PSA for 10,000 model runs are shown in the cost-effectiveness acceptability curves in *Figure 15*, and those results for no testing and laparoscopy (cut-off value of 6) are shown in *Figure 16*. These two strategies were chosen because they appear on the cost-effectiveness frontier at baseline. From *Figure 15*, no testing is always the most likely strategy to be cost-effective across willingness-to-pay values for the QALY from £0 to £200,000. In *Figure 16*, it can be seen that no testing is more likely to be cost-effective up to a willingness to pay for a QALY of approximately £150,000, at which point laparoscopy (cut-off value of 6) is more likely to be cost-effective.

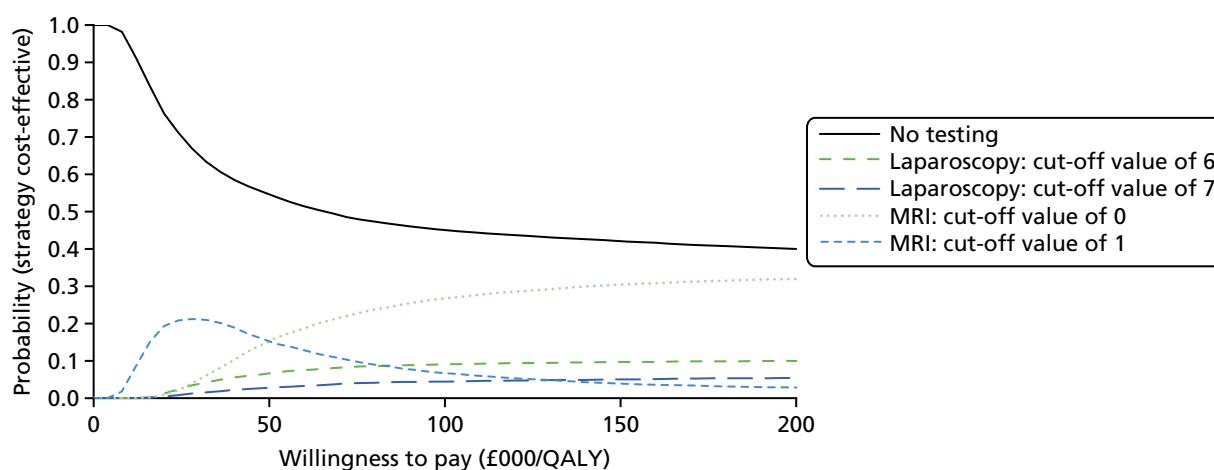


FIGURE 15 Cost-effectiveness acceptability curve showing the results of the PSA for a 6-month time horizon, for the strategies likely to be cost-effective across a range of willingness-to-pay thresholds for the QALY.

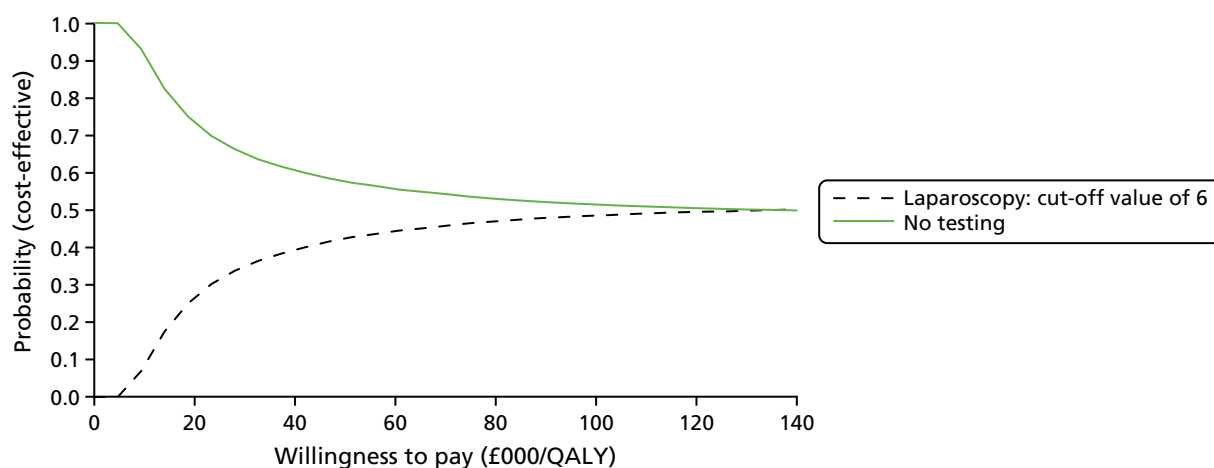


FIGURE 16 Cost-effectiveness acceptability curve showing the uncertainty in the difference between costs and effectiveness (QALYs gained) for no testing and laparoscopy (cut-off 6) for 10,000 model runs using a 6-month time horizon.

Chapter 5 Discussion

Principal findings

The diagnostic study demonstrated that, compared with the reference standard diagnosis verified at laparoscopy, MRI had high specificity but poor sensitivity for observing deep-infiltrating endometriosis (3%), endometrioma (33%), adhesions (19%) and ovarian cysts (20%). Sensitivity was higher for detecting these conditions as a cause of pelvic pain as categorised by the EIP, but not high enough to be useful. MRI correctly identified 56% of those women judged to have idiopathic CPP, but missed 46% of those women considered to have a gynaecological structural cause of CPP.

Laparoscopy was significantly more accurate than MRI in diagnosing idiopathic CPP ($p < 0.0001$), superficial peritoneal endometriosis ($p < 0.0001$), deep-infiltrating endometriosis ($p < 0.0001$) and endometrioma of the ovary ($p = 0.02$) as the cause of pelvic pain. The accuracy of laparoscopy appeared to be able to rule in these diagnoses. Laparoscopy was not able to make accurate diagnoses that adhesions were the cause of pelvic pain, nor was it able to diagnose adenomyosis.

Using MRI to identify those who require therapeutic laparoscopy would lead to 369 women in a cohort of 1000 receiving laparoscopy unnecessarily, and 136 women who required laparoscopy not receiving it. A strategy of all women receiving laparoscopy would lead to 557 women receiving unnecessary laparoscopic treatment.

Feeding these diagnostic performance data into a decision-analytic model, using a 6-month time horizon and taking the test cut-off value to inform the sensitivity and specificity values as a decision variable, it was found that diagnostic laparoscopy (with a cut-off value of 6) has an ICER value of £116,223 (cost/QALY) compared with a no-testing scenario, in which women only have a baseline examination and are then left untreated. This means that one extra QALY gained from the diagnostic laparoscopy will cost an average of £116,618 compared with no testing. This ICER is above the NICE acceptance threshold of £20,000–30,000 for the QALY. However, this value is based on a short time horizon (only 6 months) and strong assumptions, particularly regarding QALY gain in women treated with different strategies which were not actually observed in the study, which must be considered alongside these limitations, and does not provide strong enough evidence to inform clinical care.

Notably, all other scenarios, including all of the MRI scenarios across all cut-off values, were found to be either dominated (more costly and less effective) or subject to extended dominance (another strategy provides more units of benefit at a lower cost per unit of benefit). This means that MRI is not cost-effective in this setting.

As part of an extensive sensitivity analysis, the time horizon was extended to 3 years. Under these circumstances, the ICER for laparoscopy (with a cut-off value of 6) was £16,654 (cost/QALY) compared with the no-testing scenario, which would not be considered cost-effective. However, it is noted that extending the time horizon beyond the data collection period of 6 months does require further strong assumptions to be made. Nevertheless, this result does suggest that a time horizon of > 6 months should be adopted for data collection in this setting in future studies, in order for the impact of these interventions on patient QoL to be fully evaluated.

The ICER values for the outcome of cost per correct diagnosis were £579 for MRI (with a cut-off value of 11) versus no testing, £11,864 for MRI (with a cut-off value of 7) versus MRI (with a cut-off value of 11) and £13,092 for laparoscopy (with a cut-off value of 6) versus MRI (with a cut-off value of 7). In addition, the costs per patient correctly treated were £579 for MRI (with a cut-off value of 11) versus no testing, and

£9211 for laparoscopy (with a cut-off value of 6) versus MRI (with a cut-off value of 11). Although it is difficult to draw conclusions as to which strategy should be preferred based on these outcomes, as there is no acceptable threshold for either outcome, it is notable that laparoscopy (with a cut-off value of 6) continues to be the most cost-effective strategy for each.

Strengths and weaknesses of the diagnostic study

The MEDAL study examined the diagnostic value and cost-effectiveness of MRI placed before laparoscopy in a care pathway for women presenting with CPP (incorporating several target conditions) in a gynaecology clinic. If found to be accurate and cost-effective, a strategy involving use of MRI triage to direct laparoscopy could lead to more selective use of the latter, an invasive, costly test requiring use of general anaesthesia in a day-case operating theatre. Construction of an appropriate study designed to produce valid, reliable and generalisable results to guide decision-making proved to be challenging from the outset. The challenges included, among others, variation in gynaecological practice; multiplicity of target conditions with respect to cause of CPP; incongruence between observed conditions and attribution of cause to CPP symptoms; a lack of consensus on radiology standards and reporting of pelvic MRI for benign gynaecological conditions; the need to blind MRI results to avoid bias in assessment of their value; and a lack of objective reference standards for the various target conditions.

The MEDAL study was a multicentre study with recruitment from 26 gynaecology outpatient clinics in the UK. The units represented the spectrum of settings, from busy district general hospitals to specialised tertiary centres, covering the whole range of existing women's health referral networks, to hospitals with MRI facilities. These units served large, socioeconomically and ethnically diverse populations, which served to maximise the generalisability of the findings. The MEDAL study sought to approach all consecutive eligible women to obtain a suitably representative sample of women with CPP. The median interval between a MRI scan and a diagnostic laparoscopy was less than 1 month, making it unlikely that disease progression bias could influence the results, given the slow and chronic nature of the gynaecological conditions under consideration. The variation inherent in clinical and radiological practice in this mixture of settings also posed a challenge with respect to baseline data collection and MRI reporting. Establishing a standard operating procedure for MRI reporting was a key issue. We established a reporting template for CPP through a survey of an expert panel of radiologists specialising in gynaecological MRI from across the UK.⁶³ This format would facilitate the applicability of the findings of our study.

The use of laparoscopy for diagnostic evaluation in CPP was well established as the standard practice in these settings. Variations existed in the recording of history, gynaecological examination and pelvic ultrasound, before performing diagnostic laparoscopy. We standardised these during the course of the study using piloted data collections forms, which all had the input of consumers and experts to ensure sensitivity and relevance to women and their presenting complaint. These data were necessary for the care provided by the gynaecologist (who had to be kept blind to the MRI results), but their standardised collection was also necessary for replicable use by the independent gynaecologist assessing the information gained by MRI and the EIP, for establishing a reference diagnosis by consensus.

To determine the accuracy of MRI for each condition separately would require a large number of participants to accommodate the low prevalence of some conditions, such as ovarian cysts. Some pathologies are not independent of each other and could frequently be concurrently observed; for example, endometriosis can give rise to adhesions from fibrotic tissue. Furthermore, conditions observed are not always causes of the CPP symptoms. Therefore, our principal research question was developed to ascertain how MRI could be judiciously employed in a care pathway to minimise the use of laparoscopy, while maximising the correct identification of target conditions for treatments (diagnostic study) that improve QoL with the least investigative burden (decision-analytic cost-effectiveness analysis). Taking this approach, MRI test accuracy for each of the observed conditions served only as a secondary aim. Understanding the difference between diagnostic accuracy and diagnostic value in a care pathway is key to the interpretation of our findings.

To establish whether or not MRI can minimise the use of laparoscopy in a diagnostic pathway, a paired design was employed whereby both tests were performed in the study cohort. During the conduct of the study, it was necessary to blind the MRI report so that clinicians providing care remained uninfluenced by its results in their decision-making, otherwise there was a risk that the assessment of the MRI test performance in the study may be biased. The MRI test preceded laparoscopy in the study, so its findings can be interpreted for development of a triage test that can be used to direct only women with specific conditions towards laparoscopic confirmation or an operative laparoscopic procedure. By providing a MRI test prior to laparoscopy, any conditions that MRI can identify as a cause of CPP with sufficient accuracy can be treated medically without laparoscopy, or a therapeutic laparoscopy can be planned without the need for an initial diagnostic laparoscopy after the findings of the study are known. As it was feasible to perform MRI and laparoscopy in the same women with CPP in a blinded fashion, so that MRI did not interfere with laparoscopy, the paired design offered statistical efficiency over a randomised trial comparing two diagnostic strategies.

The most challenging issue was establishment of a reference diagnosis that attributed cause to the CPP symptoms. For this, we employed a panel consensus method. The panel reflected the clinical reality in which several items of information are synthesised to make a treatment decision. In this assessment, the reference standard in the MEDAL study was kept independent of the MRI test to confirm or refute the presence or absence of disease beyond reasonable doubt. A judgement was made by the expert panel about whether or not the conditions observed were a cause of the CPP symptoms without the use of MRI results, avoiding incorporation bias (except in the evaluation of adenomyosis as a target condition, as other means for this diagnosis are inadequate).

There is little evidence on how panels should be convened, how information should be presented or what the best methods for consensus are. We used a standardised approach to minimise variation in delineation of the reference standard. We established an expert panel of consultant gynaecologists who were not involved in the care of women recruited in the study. Although non-gynaecological conditions were considered as potential diagnoses, we did not include experts from other specialities, which may have contributed to the poor consensus regarding the attribution of pain to non-gynaecological causes. Using an EIP comprised solely of gynaecologists does, however, produce a diagnosis reflecting the clinical decision point, as the decision as to whether or not to perform a laparoscopy is made by gynaecologists.

Given the large amount of information to be considered per case, a piloted structured report was produced for the panel's consideration. Each panel meeting involved three experts who provided the reference diagnosis for each case presented based on the structured summary of all the data collected, with the various case report forms available in their entirety, if required. Each member individually recorded their reference diagnosis. The meeting chairperson documented the discussion that followed and how agreement was reached through consensus. Understanding the difference between the presence of a condition and a target condition being the cause of pain proved to be the main challenge. After running pilot panels, the procedures for the panel meetings were reviewed and finalised. During the course of the panel assessments, we undertook formal repeatability testing for inter-rater agreement. We employed the agreement statistics in determining the reliability of the reference standard for the interpretation of our findings. We found that agreement among experts for the assignment of gynaecological conditions as causes of CPP was either good or very good, although it was weak for non-gynaecological conditions.

The MEDAL study has been reported in accordance with the Standards for Reporting of Diagnostic Accuracy (STARD) statement. Not all items apply to the MEDAL study, as it is not a typical test-accuracy evaluation, but this statement provides for a minimum requirement to ensure completeness and transparency of reporting.

Strengths and weaknesses of the economic evaluation

The strength of this model-based analysis is that it has utilised contemporary data informing the test accuracy of diagnostic laparoscopy and MRI in their application to women with CPP. In addition, these data have provided information for the model on the impact of successfully testing and treating women with structural causes of CPP on their QoL. Moreover, as part of this analysis, the cut-off value on the ROC curve to inform the sensitivity and specificity of each test was treated as a decision variable, meaning that the most cost-effective test at its optimum cut-off value could be identified.

This analysis has a number of weaknesses that need to be acknowledged. QoL data were collected from women over a 6-month period only, meaning that a time horizon of only 6 months could be adopted in the model, unless strong assumptions were made about the QoL and costs incurred beyond 6 months. During the sensitivity analysis, the time horizon was extended up to 3 years, but this relied on the assumptions that QoL did not change beyond 6 months and no costs were incurred by either strategy beyond that time. It is acknowledged that CPP among women is a problem that can often take many medical consultations to resolve, and so it is likely that the positive benefits of properly diagnosing and treating a structural cause of CPP will be felt long after the 6-month time horizon considered here.

Given that this model utilised data from a test-accuracy study in which all women received both tests, it was difficult to tease out the impact of the different test results and the resulting treatment on patient QoL. This meant that assumptions had to be made regarding how negative test results might affect women. This is really an unavoidable issue with model-based analyses that examine medical tests, but is acknowledged here nonetheless.

This economic analysis has considered only the diagnosis and treatment of structural causes of CPP, and has not included other possible causes of CPP, such as PID. Including more possible causes of CPP would make the model more cumbersome, and would also require the different treatment and test characteristics for the new causes to be incorporated into the model. This means that the full patient benefit of laparoscopy and MRI has not been considered here. Nevertheless, varying the prevalence of structural causes of CPP was undertaken in the sensitivity analysis presented here, and slightly increasing the prevalence could be regarded as a proxy for the inclusion of other causes of CPP, and thus this does help to show how their inclusion might impact on the conclusions drawn from the model.

In this model-based economic evaluation, a decision tree was utilised rather than a Markov model. This was attributable to the short time horizon and the fact that women acquiring new cases of CPP over time were not considered in the study. However, it is acknowledged that an individual sampling model could have been used, whereby the different structural causes of CPP would be described for each patient. However, given that the testing and treatment pathways for many of the causes of CPP considered in this study are the same, this approach was considered unnecessary as these pathways could easily be described using a decision tree.

Public and patient involvement

We have been supported throughout the project by the Pelvic Pain Support Network (PPSN) and, in particular, its chairperson. Public and patient involvement was crucial in improving the acceptability of the MEDAL study and promoting recruitment. We engaged with the PPSN chairperson throughout, developing an appreciation of the apprehension and uncertainty surrounding MRI and laparoscopy among women, as well as the opinions of clinicians and barriers to accessing MRI that the chairperson had encountered. This prompted us to provide a study-specific MRI information leaflet to potential participants. We also believed that it was important to establish a thorough understanding of all aspects of CPP and the potential underlying conditions, and so we designed the initial patient questionnaire with input from the PPSN.

We will engage with the PPSN regarding the dissemination of our findings, providing a plain English summary of the findings and the uncertainties around the evidence we have discussed here. This will be distributed via the PPSN website and e-newsletter. Any future research groups taking forward the research recommendations from this project would benefit from engaging with the PPSN.

Interpretation of the findings

To our knowledge, this is the first study that has examined the diagnostic value and cost-effectiveness of MRI compared with diagnostic laparoscopy in women with CPP.

The risks of MRI revolve around changing radiofrequencies and magnetic fields, which can theoretically produce heat that is absorbed by the body tissue; this is not known to produce any side effects. Conversely, like any surgery, a laparoscopy is not without its risks. It involves day care admission and general anaesthesia. Laparoscopy involves minimal invasion of body tissues, but its complications can include damage to organs inside the abdomen and wound infections. Therefore, if MRI could minimise the use of laparoscopy, it could be beneficial to women and health services. Our diagnostic study found that MRI tests were not accurate for detecting structural gynaecological conditions. Furthermore, it did not add value over and above baseline history, examination and pelvic ultrasound in verifying the absence of a cause of CPP symptoms (before performing diagnostic laparoscopy).

With respect to specific gynaecological conditions, MRI was most accurate for endometrioma, adhesions and fibroids when compared with observations at laparoscopy but, even so, its performance was too inadequate to be useful. Evaluation of non-gynaecological conditions suffered from poverty of agreement over verification of the diagnosis by an expert panel, and this limits the useful interpretation of these data.

The health economic analysis found that MRI was never more cost-effective than laparoscopy. Given the poor accuracy of MRI, this conclusion was expected. Our extended analysis comparing laparoscopy and MRI with a theoretical no-testing scenario revealed the interesting observation that, at the 6-month follow-up, a strategy of no laparoscopy or MRI would be the most cost-effective – this finding is likely to be strongly affected by the short follow-up period (the benefits of successful treatment of CPP are likely to last many years), and the assumptions made about QoL without treatment (which could not be directly observed in the study).

Implications for practice

Magnetic resonance imaging was dominated by laparoscopy in the differential diagnosis of women presenting to gynaecology clinics with CPP. It did not add value to the information, already gained from history, examination and ultrasound, about idiopathic CPP and various gynaecological conditions before considering a laparoscopy. The little value it did show for a few conditions, such as endometrioma, could be obtained only at the cost of undertaking large numbers of scans, and was not cost-effective, as the small diagnostic gain came tagged with additional costs.

Unanswered questions and further research

The evaluation of subjective tests for differential diagnosis in care pathways targeting various conditions for treatment requires methodological development. Further data on the individual and joint distributions of the prevalence of causes of pain would be beneficial. In CPP, the development of diagnostic prediction models incorporating biomarkers into baseline data may improve the use of diagnostic laparoscopy; however, research will first be required to determine objective diagnostic criteria for the various target conditions and valid biomarkers for endometriosis.

Acknowledgements

The Magnetic resonance imaging for Establishing Diagnosis Against Laparoscopy study group

The MEDAL study collaborators (*principal investigators/investigators/research nurses*).

Basildon University Hospital

Moboladale Ojutiku, Andrew Hails, Sami Khan, Kerry Goodsell, Emily Redman, Stacey Pepper and Marco Bondoc.

Birmingham Women's Hospital

Moji Balogun, Elizabeth Ewers, Aamir Khan, Tina Verghuse, Karen Davies, Amy Wilbraham, Chloe O'Hara, Natalie Cooka, Natalie Nunes, Yousri Afifi and Pallavi Latthe.

Chelsea and Westminster

Robert Richardson, Amer Raza, Priya Narayanan, Rhian Bull, Kylie Norrie, Sarah Ladd and Paul Cofie.

Cumberland Infirmary

Nalini Munjuluri, Toni Wilson, Job Berry, Rachel England, Linda Telford and Alexa Lilley.

Dorset County Hospital

Mamdouh Shoukrey, Karen Hogben, Andrew Gibbins, Katheryn Lawrence and Sue O'Kerwin.

Royal Infirmary of Edinburgh

Andrew Horne, Ann Doust, Scott Semple, Graham McKillop, Hayley Cuthbert, Pamela Haig, Lisa Derr, Fiona Steel and Helen Dewart.

Furness General Hospital

Vincent Bamigboye, Julia Alcide, Ali Ansarai, Amoah Christian, Burns Karen and Sanjay Sinha.

Homerton University Hospital

Sandra Watson, Akosua Aboagaye, Emma Goldstraw and Nicola Jenkerson.

Liverpool Women's Hospital

Adel Soltan, Amy Smith, Gillian Smith, Kathie Cooke and Lesley Harris.

Medway Maritime Hospital

Sadoon Sadoon, Abu Ahmed, Dorothy Chakani, Emily Tam, Eunice Emeakaraoha, John Duckett, Jonathan Duckett, Ibrahim Ahmed, Pam Arkhurst and Sarah Burrows.

Musgrove Park Hospital

Guy Fender, Korinna Andrews, Paul Burn, Ann England and Michelle Farrar.

Newham Hospital

Antonios Antoniou, Joanne Hutchinson and Rashna Chenoy.

Ninewells Hospital

Refaat Youssef, Alison Boyle, Audrey Lyall, Deborah Forbes, Pamela Duthie, Audrey Lyall, Thiru Sudarshan, Jennifer Brady and John Brunton.

North Devon District Hospital

Seumas Eckford, Amanda Skinner, Colin Barratt, Emad Megaly, Osama Eskandar, Fiona Hammonds, Geraldine Belcher, Helen Shelton, Afaf Diyaf, Jennifer Macpherson and Natalie Taylor.

Nottingham University Hospital

Nick Raine-Fenning, Ann Selby, Anna Molnar, Corah Ohadike, Maruti Kumaran, Elizabeth Barnes, Gill Kirkwood, Kannamannadiar Jayaprakasan, Martin Powell and Maruti Kimaran.

Royal Hallamshire Hospital

Mary Connor, Andrew Baxter, Chris Wragg, Clare Pye, Sheila Duffy, Julie Metherall, Mary Connor, Shahram Abdi and Sheila Duffy.

The Royal London Hospital

Elizabeth Ball, Andrea Rockall, Seema Anoushka Tirlapur, Mathew Hogg, Mun Lim, Natasha Kaba, Shoreh Beksi and Teresita Beeston.

Royal Preston Hospital

Sanjeev Prashar, Sajata Gupta, Patrick Keating, Sujata Gupta, Mike Dobson, Khalil Abdo, Himansuskhar Mallik, Sean Hughes, Pauline McDonald and Julie Butler.

South Tyneside District Hospital

Steve Orife, Oliver Schulte, Richard Cooper and Linda McNamee.

Southend University Hospital

Justin Winston, Venkat Raman, Cheng Lee, Gemma Ogden, Maryam Zare, Jane Hollywood, Onie Hove, Narayanaswamy Venkat Raman, Sid Liyanage and Neil Quinn.

Stafford Hospital

Kirk Chin, Elizabeth Gunning, Tracey Harrison, Susan Hendy, Lynne Ball, Dawn Sirdefield and Jill Stacey.

Sunderland Royal Hospital

Menem Yossry, Judith Ormonde, Karen Armstrong, Jemma Fenwick, Kim Hinshaw, Ahma Ahmed, Aarti Ulal, Gill Campbell, Denise Milford, Eileen Walton, Lesley Hewitt and Jemma Fenwick.

University Hospital Crosshouse

Inna Sokolova, Danielle Gilmour, Karen Gray and Margo Henry.

University Hospital of North Durham

Partha Sengupta, Chooi May Lee, Jean Dent and Partha Sengupta.

University Hospital of North Staffordshire

Jason Cooper, Anne Harrison, Bruce Jarvest, Chan, Junny, Cooper Jason, El-Gizawy Zeiad, Gourab Misra, Emma Hubball, Suzanne Jerreat, Junny Chan, Joanne Lakin, O'Brien Shaughn, Robha Ramakoti and Amanda Redford.

Whipps Cross University Hospital

Funlayo Odejinmi, Zandile Maseko and Oladimeji Olowu.

Expert independent panel (for establishing a consensus reference diagnosis)

Mr Christian Becker, John Radcliffe Hospital, Oxford; Mr Alfred Cutner, University College Hospital, London; Mr Patrick Chien (Expert Panel Chairperson), Ninewells Hospital, Dundee; Mr Justin Clark, Birmingham Women's Hospital, Birmingham; Mr Yemi Coker, Queen's Hospital, Romford; Mr Shane Duffy, Chelsea and Westminster Hospital, London; Mr Jed Hawe, Countess of Chester Hospital, Chester; Mr Kevin Phillips, Castle Hill Hospital, Cottingham; Dr Caroline Overton, University Hospitals Bristol NHS Foundation Trust, Bristol; Mr Adam Rosenthal, St Bartholomew's Hospital, London; Mr Ertan Saridogan, University College Hospital, London; Mr Tom Smith Walker, Royal Cornwall Hospital, Truro; Mr Jim Thornton, Nottingham City Hospital, Nottingham; Mr Sanjay Vyas, Spire Bristol Hospital, Bristol; and Miss Ephie Yasmin, University College London Hospital, London.

Independent radiology review committee (for establishing a consensus diagnosis from the magnetic resonance imaging scan)

Dr Mark Blakeman, New Cross Hospital, Wolverhampton; Dr Claire Keaney, Sandwell General Hospital, West Bromwich/Birmingham City Hospital, Birmingham; and Dr Carolina Lopez, Bedford Hospital, Bedford.

The Study Steering and Data Monitoring Committees

The SSC provided independent supervision for the study, providing advice to the investigators and the sponsor on all aspects of the study and affording protection for women by ensuring that the study was conducted as applicable to the MRC guidelines for good clinical practice in clinical trials.

For the purposes of this study, the SSC nominated and convened a three-member independent Data Monitoring and Ethics Committee (DMEC) from within its membership, which did not include study researchers:

- Professor Ovrang Djahanbakhch (SSC Chairperson and Head of the Unit for Women's Health, The Royal London Hospital, London)
- Professor Ben Willem Mol (DMEC Chairperson and Professor of Obstetrics and Gynaecology, University of Adelaide, Adelaide, SA, Australia)
- Professor Javier Zamora (until 2013, Senior Lecturer, Barts and The London School of Medicine and Dentistry, London)
- Dr Teresa Perez (replaced Dr Zamora in 2013; Complutense University of Madrid, Madrid, Spain)
- Dr George Harrison (Consultant in Pain, Queen Elizabeth Hospital Birmingham, Birmingham, UK)
- Ms Judy Birch (Chief Executive, Pelvic Pain Support Network, Poole, UK).

Study management group

The MEDAL study was co-ordinated by the Birmingham Clinical Trials Unit at the University of Birmingham and we acknowledge all of the hard work of all of the staff involved in the study, including Lee Priest (Trial Co-ordinator) and Julia Seeley (Data Manager). We thank Neil Winkles for designing and developing the study database. We thank Rita Champaneria and Bella Harris for the systematic review and reference support. We would also like to thank Christopher Webb for the organisation of the study meetings. The MEDAL study sponsor organisation was the Queen Mary University of London, and we acknowledge Kathline Mulligan (Senior Clinical Trials Manager until 2012), Julie Dodds (Senior Clinical Trials Manager from 2012), Tracy Holtham and Camille Paulsen for their support.

The study management group comprised Khalid Khan (Queen Mary University of London); Elizabeth Ball (The Royal London Hospital); Jon Deeks (University of Birmingham); Jane Daniels (University of Birmingham); Judy Birch (Pelvic Pain Support Network); Lee Middleton (University of Birmingham); Lee Priest (University of Birmingham); Julia Seeley (University of Birmingham); Seema Anushka Tirlapur (Queen Mary University of London); Julie Dodds (Queen Mary, University of London); and Teresita Beeston (Queen Mary University of London).

We are grateful to the British Society of Gynaecological Endoscopists.

Research governance

The study was conducted in accordance with the principles of MRC Guidelines for Good Clinical Practice in Clinical Trials (1998) and the appropriate NHS Research Governance Frameworks.

All centres will be required to sign an Investigator's Agreement, detailing their commitment to accrual, compliance, good clinical practice, confidentiality and publication. Deviations from the agreement will be monitored and the SSC will decide if any action needs to be taken, such as withdrawal of funding or suspension of the centre. Proof of training in the principles of good clinical practice and informed consent may be required.

Study advisors

Dr Pallavi Latthe (Consultant Obstetrician and Gynaecologist, Birmingham Women's Hospital) was a recruiting investigator and helped to develop the working diagnosis form.

Dr Susan Mallet (Senior Lecturer in Medical Statistics in the Test Evaluation Research Group, University of Birmingham) advised on the analysis of the study.

Dr Nigel Cowan (Consultant Clinical Radiologist, Nuffield Health Oxford Hospital) provided guidance on the criteria for the assessment of the target conditions investigated in the study.

Professor Charles Knowles (Clinical Professor of Surgical Research, Barts and the London School of Medicine and Dentistry Unit) provided guidance on bowel symptoms.

Professor Kasim Aziz (Professor of Neurogastroenterology, Queen Mary University of London) provided guidance on bowel symptoms.

Dr Jonathan Hill (Keele University) provided guidance on physiotherapy and musculoskeletal aspects.

Finally, we would like to thank all of the women who consented to participate in this study. The study would not have been possible without them.

Contributions of authors

Professor Khalid S Khan (Consultant Gynaecologist and lead applicant) conceived the idea for the study, contributed to the design, delivery, analysis and interpretation, recruited women, and, as the chief investigator, had overall responsibility for the MEDAL study.

Mr Konstantinos Tryposkiadis (Statistician) performed the analyses for the accuracy study and contributed to the interpretation and drafting of the report.

Dr Seema A Tirlapur (Clinical Research Fellow) recruited participants, developed the process for the EIP and contributed to the delivery of the accuracy study and the development of the economic model.

Mr Lee J Middleton (Senior Statistician) contributed to the design, analysis and interpretation of the accuracy study.

Dr Andrew J Sutton (Senior Economic Research Fellow) contributed to the design and interpretation of the economic evaluation, performed the analysis and drafted the economic study chapter (see *Chapter 4*).

Mr Lee Priest (Senior Trial Co-ordinator) was responsible for the day-to-day management and delivery of all study components, and developed the process for the EIP panel and edited the report.

Dr Elizabeth Ball (Consultant Gynaecologist and co-applicant) contributed to the design and delivery of the accuracy study and the development of the economic model, and recruited participants.

Dr Moji Balogun (Consultant Radiologist and co-applicant) contributed to the design of the accuracy study and the development of the MRI reporting standards.

Dr Anju Sahdev (Consultant Radiologist) contributed to the development of the MRI reporting standards.

Professor Tracy Roberts (Professor of Health Economics and co-applicant) contributed to the design, analysis and interpretation of the economic evaluation.

Mrs Judy Birch (Chairperson, Pelvic Pain Support Network and co-applicant) contributed to the design of the accuracy study and commented on the acceptability of participant-facing material.

Dr Jane P Daniels (Deputy Director of Birmingham Clinical Trials Unit and co-applicant) contributed to the design, delivery and interpretation of the MEDAL study, the first draft and overall editing of the final report.

Professor Jonathan J Deeks (Professor of Medical Statistics and co-applicant) defined the design and statistical analysis of the accuracy study and contributed to the interpretation of the accuracy study and economic model.

Publications

Tirlapur SA, Priest L, Daniels JP, Khan KS on behalf of the MEDAL study management group. How do we define the term idiopathic? *Curr Opin Obstet Gynecol* 2013;**25**:468–7.

Tirlapur SA, Priest L, Wojdyla D, Khan KS on behalf of the MEDAL Study. Bladder pain syndrome: validation of simple tests for diagnosis in women with chronic pelvic pain: BRaVADO study protocol. *Reprod Health* 2013;**10**:61.

Tirlapur SA, Daniels JP, Khan K on behalf of the MEDAL Study. Chronic pelvic pain: how does noninvasive imaging compare with diagnostic laparoscopy? *Curr Opin Obstet Gynecol* 2015;**27**:445–8.

Bharwani N, Tirlapur SA, Balogun M, Priest L, Khan KS, Zamora J, Sadhev A. MRI reporting standard for chronic pelvic pain: consensus development. *Br J Radiol* 2016;**89**:20140615.

Data sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to available anonymised data may be granted following review.

Patient data

This work uses data provided by women and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: <https://understandingpatientdata.org.uk/data-citation>.

References

1. Williams RE, Hartmann KE, Steege JF. Documenting the current definitions of chronic pelvic pain: implications for research. *Obstet Gynecol* 2004;**103**:686–91. <https://doi.org/10.1097/01.AOG.0000115513.92318.b7>
2. Tirlapur SA, Priest L, Daniels JP, Khan KS, MEDAL Study Management Group. How do we define the term idiopathic? *Curr Opin Obstet Gynecol* 2013;**25**:468–73. <https://doi.org/10.1097/GCO.0000000000000025>
3. Daniels JP, Khan KS. Chronic pelvic pain in women. *BMJ* 2010;**341**:c4834. <https://doi.org/10.1136/bmj.c4834>
4. Ahangari A. Prevalence of chronic pelvic pain among women: an updated review. *Pain Physician* 2014;**17**:E141–7.
5. Zondervan KT, Yudkin PL, Vessey MP, Dawes MG, Barlow DH, Kennedy SH. Prevalence and incidence of chronic pelvic pain in primary care: evidence from a national general practice database. *Br J Obstet Gynaecol* 1999;**106**:1149–55. <https://doi.org/10.1111/j.1471-0528.1999.tb08140.x>
6. Zondervan K, Barlow DH. Epidemiology of chronic pelvic pain. *Baillieres Best Pract Res Clin Obstet Gynaecol* 2000;**14**:403–14. <https://doi.org/10.1053/beog.1999.0083>
7. Henzl MR. Dysmenorrhoea; achievements and challenge. *Sex Med Today* 1985;**9**:7–11.
8. Howard FM. The role of laparoscopy in chronic pelvic pain: promise and pitfalls. *Obstet Gynecol Surv* 1993;**48**:357–87. <https://doi.org/10.1097/00006254-199306000-00001>
9. Zondervan KT, Yudkin PL, Vessey MP, Jenkinson CP, Dawes MG, Barlow DH, Kennedy SH. The community prevalence of chronic pelvic pain in women and associated illness behaviour. *Br J Gen Pract* 2001;**51**:541–7.
10. Grace V, Zondervan K. Chronic pelvic pain in women in New Zealand: comparative well-being, comorbidity, and impact on work and other activities. *Health Care Women Int* 2006;**27**:585–99. <https://doi.org/10.1080/07399330600803725>
11. Aslam N, Harrison G, Khan K, Patwardhan S. Visceral hyperalgesia in chronic pelvic pain. *BJOG* 2009;**116**:1551–5. <https://doi.org/10.1111/j.1471-0528.2009.02305.x>
12. Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Hum Reprod* 2012;**27**:1292–9. <https://doi.org/10.1093/humrep/des073>
13. Baranowski A, Abrams P, Berger RE. *Taxonomy of Pelvic Pain. Classification of Chronic Pain*. 2nd edn. Washington, DC: International Association for the Study of Pain; 2012.
14. Champaneria R, Shah L, Moss J, Gupta JK, Birch J, Middleton LJ, Daniels JP. The relationship between pelvic vein incompetence and chronic pelvic pain in women: systematic reviews of diagnosis and treatment effectiveness. *Health Technol Assess* 2016;**20**(5). <https://doi.org/10.3310/hta20050>
15. Gruppo Italiano per lo Studio dell'Endometriosi. Relationship between stage, site and morphological characteristics of pelvic endometriosis and pain. *Hum Reprod* 2001;**16**:2668–71. <https://doi.org/10.1093/humrep/16.12.2668>
16. Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. *Neuron* 2007;**55**:377–91. <https://doi.org/10.1016/j.neuron.2007.07.012>

17. Daniels J, Gray R, Hills RK, Latthe P, Buckley L, Gupta J, *et al.* Laparoscopic uterosacral nerve ablation for alleviating chronic pelvic pain: a randomized controlled trial. *JAMA* 2009;**302**:955–61. <https://doi.org/10.1001/jama.2009.1268>
18. Fall M, Baranowski AP, Elneil S, Engeler D, Hughes J, Messelink EJ, *et al.* EAU guidelines on chronic pelvic pain. *Eur Urol* 2010;**57**:35–48. <https://doi.org/10.1016/j.eururo.2009.08.020>
19. Ballard K, Lowton K, Wright J. What's the delay? A qualitative study of women's experiences of reaching a diagnosis of endometriosis. *Fertil Steril* 2006;**86**:1296–301. <https://doi.org/10.1016/j.fertnstert.2006.04.054>
20. Royal College of Obstetricians and Gynaecologists. *The Initial Management of Chronic Pelvic Pain: Green-top Guideline No. 41*. London: Royal College of Obstetricians and Gynaecologists; 2005.
21. Hadfield R, Mardon H, Barlow D, Kennedy S. Delay in the diagnosis of endometriosis: a survey of women from the USA and the UK. *Hum Reprod* 1996;**11**:878–80. <https://doi.org/10.1093/oxfordjournals.humrep.a019270>
22. Matsuzaki S, Canis M, Pouly JL, Rabischong B, Botchorishvili R, Mage G. Relationship between delay of surgical diagnosis and severity of disease in patients with symptomatic deep infiltrating endometriosis. *Fertil Steril* 2006;**86**:1314–16. <https://doi.org/10.1016/j.fertnstert.2006.03.048>
23. Prentice A. Medical management of chronic pelvic pain. *Baillieres Best Pract Res Clin Obstet Gynaecol* 2000;**14**:495–9. <https://doi.org/10.1053/beog.1999.0087>
24. Okaro E, Condous G, Khalid A, Timmerman D, Ameye L, Huffel SV, Bourne T. The use of ultrasound-based 'soft markers' for the prediction of pelvic pathology in women with chronic pelvic pain – can we reduce the need for laparoscopy? *BJOG* 2006;**113**:251–6. <https://doi.org/10.1111/j.1471-0528.2006.00849.x>
25. Cervigni M, Natale F. Gynecological disorders in bladder pain syndrome/interstitial cystitis patients. *Int J Urol* 2014;**21**(Suppl. 1):85–8. <https://doi.org/10.1111/iju.12379>
26. Walker EA, Gelfand AN, Gelfand MD, Green C, Katon WJ. Chronic pelvic pain and gynecological symptoms in women with irritable bowel syndrome. *J Psychosom Obstet Gynaecol* 1996;**17**:39–46. <https://doi.org/10.3109/01674829609025662>
27. Longstreth GF. Irritable bowel syndrome and chronic pelvic pain. *Obstet Gynecol Surv* 1994;**49**:505–7. <https://doi.org/10.1097/00006254-199407000-00027>
28. Thornton JG, Morley S, Lilleyman J, Onwude JL, Currie I, Crompton AC. The relationship between laparoscopic disease, pelvic pain and infertility; an unbiased assessment. *Eur J Obstet Gynecol Reprod Biol* 1997;**74**:57–62. [https://doi.org/10.1016/S0301-2115\(97\)00082-1](https://doi.org/10.1016/S0301-2115(97)00082-1)
29. Fauconnier A, Chapron C. Endometriosis and pelvic pain: epidemiological evidence of the relationship and implications. *Hum Reprod Update* 2005;**11**:595–606. <https://doi.org/10.1093/humupd/dmi029>
30. Nisenblatt V, Prentice L, Bossuyt PM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database Syst Rev* 2016;**7**:CD012281. <https://doi.org/10.1002/14651858.CD012281>
31. Howard FM. The role of laparoscopy as a diagnostic tool in chronic pelvic pain. *Baillieres Best Pract Res Clin Obstet Gynaecol* 2000;**14**:467–94. <https://doi.org/10.1053/beog.1999.0086>
32. Howard FM. Laparoscopic evaluation and treatment of women with chronic pelvic pain. *J Am Assoc Gynecol Laparosc* 1994;**1**:325–31. [https://doi.org/10.1016/S1074-3804\(05\)80797-2](https://doi.org/10.1016/S1074-3804(05)80797-2)
33. Khan KN, Fujishita A, Kitajima M, Hiraki K, Nakashima M, Masuzaki H. Occult microscopic endometriosis: undetectable by laparoscopy in normal peritoneum. *Hum Reprod* 2014;**29**:462–72. <https://doi.org/10.1093/humrep/det438>

34. Moore J, Ziebland S, Kennedy S. 'People sometimes react funny if they're not told enough': women's views about the risks of diagnostic laparoscopy. *Health Expect* 2002;**5**:302–9. <https://doi.org/10.1046/j.1369-6513.2002.00192.x>
35. McGowan L, Luker K, Creed F, Chew-Graham CA. How do you explain a pain that can't be seen? The narratives of women with chronic pelvic pain and their disengagement with the diagnostic cycle. *Br J Health Psychol* 2007;**12**:261–74. <https://doi.org/10.1348/135910706X104076>
36. Walter AJ, Hentz JG, Magtibay PM, Cornella JL, Magrina JF. Endometriosis: correlation between histologic and visual findings at laparoscopy. *Am J Obstet Gynecol* 2001;**184**:1407–11. <https://doi.org/10.1067/mob.2001.115747>
37. Wykes CB, Clark TJ, Khan KS. Accuracy of laparoscopy in the diagnosis of endometriosis: a systematic quantitative review. *BJOG* 2004;**111**:1204–12. <https://doi.org/10.1111/j.1471-0528.2004.00433.x>
38. Stratton P, Winkel C, Premkumar A, Chow C, Wilson J, Hearn-Stokes R, et al. Diagnostic accuracy of laparoscopy, magnetic resonance imaging, and histopathologic examination for the detection of endometriosis. *Fertil Steril* 2003;**79**:1078–85. [https://doi.org/10.1016/S0015-0282\(03\)00155-9](https://doi.org/10.1016/S0015-0282(03)00155-9)
39. Bazot M, Bornier C, Dubernard G, Roseau G, Cortez A, Daraï E. Accuracy of magnetic resonance imaging and rectal endoscopic sonography for the prediction of location of deep pelvic endometriosis. *Hum Reprod* 2007;**22**:1457–63. <https://doi.org/10.1093/humrep/dem008>
40. Katayama M, Masui T, Kobayashi S, Ito T, Sakahara H, Nozaki A, Kabasawa H. Evaluation of pelvic adhesions using multiphase and multislice MR imaging with kinematic display. *AJR Am J Roentgenol* 2001;**177**:107–10. <https://doi.org/10.2214/ajr.177.1.1770107>
41. Tukeva TA, Aronen HJ, Karjalainen PT, Molander P, Paavonen T, Paavonen J. MR imaging in pelvic inflammatory disease: comparison with laparoscopy and US. *Radiology* 1999;**210**:209–16. <https://doi.org/10.1148/radiology.210.1.r99ja04209>
42. Munday PE. Pelvic inflammatory disease – an evidence-based approach to diagnosis. *J Infect* 2000;**40**:31–41. <https://doi.org/10.1053/jinf.1999.0609>
43. Schindlbeck C, Dziura D, Mylonas I. Diagnosis of pelvic inflammatory disease (PID): intra-operative findings and comparison of vaginal and intra-abdominal cultures. *Arch Gynecol Obstet* 2014;**289**:1263–9. <https://doi.org/10.1007/s00404-014-3150-7>
44. Champaneria R, Abedin P, Daniels J, Balogun M, Khan KS. Ultrasound scan and magnetic resonance imaging for the diagnosis of adenomyosis: systematic review comparing test accuracy. *Acta Obstet Gynecol Scand* 2010;**89**:1374–84. <https://doi.org/10.3109/00016349.2010.512061>
45. Hsu AL, Khachikyan I, Stratton P. Invasive and non-invasive methods for the diagnosis of endometriosis. *Clin Obstet Gynecol* 2010;**53**:413–19.
46. Dueholm M, Lundorf E. Transvaginal ultrasound or MRI for diagnosis of adenomyosis. *Curr Opin Obstet Gynecol* 2007;**19**:505–12. <https://doi.org/10.1097/GCO.0b013e3282f1bf00>
47. Dueholm M, Lundorf E, Hansen ES, Ledertoug S, Olesen F. Accuracy of magnetic resonance imaging and transvaginal ultrasonography in the diagnosis, mapping, and measurement of uterine myomas. *Am J Obstet Gynecol* 2002;**186**:409–15. <https://doi.org/10.1067/mob.2002.121725>
48. Bazot M, Bharwani N, Huchon C, Kinkel K, Cunha TM, Guerra A, et al. European Society of Urogenital Radiology (ESUR) guidelines: MR imaging of pelvic endometriosis. *Eur Radiol* 2017;**27**:2765–2775. <https://doi.org/10.1007/s00330-016-4673-z>
49. Younas K, Majoko F, Sheard K, Edwards C, Bunkheila A. Select and treat at laparoscopy and dye test improves the spontaneous pregnancy. *Hum Fertil* 2014;**17**:56–9. <https://doi.org/10.3109/14647273.2014.880522>

50. Ball E, Koh C, Janik G, Davis C. Gynaecological laparoscopy: 'see and treat' should be the gold standard. *Curr Opin Obstet Gynecol* 2008;**20**:325–30. <https://doi.org/10.1097/GCO.0b013e32830002bb>
51. Cheong YC, Reading I, Bailey S, Sadek K, Ledger W, Li TC. Should women with chronic pelvic pain have adhesiolysis? *BMC Womens Health* 2014;**14**:36. <https://doi.org/10.1186/1472-6874-14-36>
52. Duffy JM, Arambage K, Correa FJ, Olive D, Farquhar C, Garry R, et al. Laparoscopic surgery for endometriosis. *Cochrane Database Syst Rev* 2014;**3**:CD011031. <https://doi.org/10.1002/14651858.CD011031>
53. Farquhar C, Sutton C. The evidence for the management of endometriosis. *Curr Opin Obstet Gynecol* 1998;**10**:321–32. <https://doi.org/10.1097/00001703-199808000-00007>
54. Bhavne Chittawar P, Franik S, Pouwer AW, Farquhar C. Minimally invasive surgical techniques versus open myomectomy for uterine fibroids. *Cochrane Database Syst Rev* 2014;**10**:CD004638. <https://doi.org/10.1002/14651858.CD004638.pub3>
55. Wu L, Wu Q, Liu L. Oral contraceptive pills for endometriosis after conservative surgery: a systematic review and meta-analysis. *Gynecol Endocrinol* 2013;**29**:883–90. <https://doi.org/10.3109/09513590.2013.819085>
56. Abou-Setta AM, Houston B, Al-Inany HG, Farquhar C. Levonorgestrel-releasing intrauterine device (LNG-IUD) for symptomatic endometriosis following surgery. *Cochrane Database Syst Rev* 2013;**1**:CD005072. <https://doi.org/10.1002/14651858.CD005072.pub3>
57. Wong AY, Tang LC, Chin RK. Levonorgestrel-releasing intrauterine system (Mirena) and Depot medroxyprogesterone acetate (Depoprovera) as long-term maintenance therapy for patients with moderate and severe endometriosis: a randomised controlled trial. *Aust N Z J Obstet Gynaecol* 2010;**50**:273–9. <https://doi.org/10.1111/j.1479-828X.2010.01152.x>
58. Lewis SC, Bhattacharya S, Wu O, Vincent K, Jack SA, Critchley HO, et al. Gabapentin for the management of chronic pelvic pain in women (GaPP1): a pilot randomised controlled trial. *PLOS ONE* 2016;**11**:e0153037. <https://doi.org/10.1371/journal.pone.0153037>
59. Jansen FW, Kapiteyn K, Trimbos-Kemper T, Hermans J, Trimbos JB. Complications of laparoscopy: a prospective multicentre observational study. *Br J Obstet Gynaecol* 1997;**104**:595–600. <https://doi.org/10.1111/j.1471-0528.1997.tb11539.x>
60. Kontoravdis A, Chryssikopoulos A, Hassiakos D, Liapis A, Zourlas PA. The diagnostic value of laparoscopy in 2365 patients with acute and chronic pelvic pain. *Int J Gynaecol Obstet* 1996;**52**:243–8. [https://doi.org/10.1016/0020-7292\(95\)02611-8](https://doi.org/10.1016/0020-7292(95)02611-8)
61. Chapron C, Querleu D, Bruhat MA, Madelenat P, Fernandez H, Pierre F, Dubuisson JB. Surgical complications of diagnostic and operative gynaecological laparoscopy: a series of 29,966 cases. *Hum Reprod* 1998;**13**:867–72. <https://doi.org/10.1093/humrep/13.4.867>
62. Physician Insurers Association of America. *Laparoscopic Injury Study*. Rockville, MD: Physician Insurers Association of America; 2000.
63. Bharwani N, Tirlapur SA, Balogun M, Priest L, Khan KS, Zamora J, Sahdev A. MRI reporting standard for chronic pelvic pain: consensus development. *Br J Radiol* 2016;**89**:20140615. <https://doi.org/10.1259/bjr.20140615>
64. Honest H, Khan KS. Reporting of measures of accuracy in systematic reviews of diagnostic literature. *BMC Health Serv Res* 2002;**2**:4. <https://doi.org/10.1186/1472-6963-2-4>
65. Bazot M, Lafont C, Rouzier R, Roseau G, Thomassin-Naggara I, Daraï E. Diagnostic accuracy of physical examination, transvaginal sonography, rectal endoscopic sonography, and magnetic resonance imaging to diagnose deep infiltrating endometriosis. *Fertil Steril* 2009;**92**:1825–33. <https://doi.org/10.1016/j.fertnstert.2008.09.005>

66. Chamié LP, Blasbalg R, Gonçalves MO, Carvalho FM, Abrão MS, de Oliveira IS. Accuracy of magnetic resonance imaging for diagnosis and preoperative assessment of deeply infiltrating endometriosis. *Int J Gynaecol Obstet* 2009;**106**:198–201. <https://doi.org/10.1016/j.ijgo.2009.04.013>
67. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;**11**(50). <https://doi.org/10.3310/hta11500>
68. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002;**324**:669–71. <https://doi.org/10.1136/bmj.324.7338.669>
69. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006;**332**:1089–92. <https://doi.org/10.1136/bmj.332.7549.1089>
70. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;**11**:88–94. <https://doi.org/10.1177/0272989X9101100203>
71. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837–45. <https://doi.org/10.2307/2531595>
72. Fleiss J. *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley & Sons, Inc.; 1981.
73. Walker JJ, Irvine G. How should we approach the management of pelvic pain? *Gynecol Obstet Invest* 1998;**45**(Suppl. 1):6–10. <https://doi.org/10.1159/000052846>
74. Melzack R. The short-form McGill Pain Questionnaire. *Pain* 1987;**30**:191–7. [https://doi.org/10.1016/0304-3959\(87\)91074-8](https://doi.org/10.1016/0304-3959(87)91074-8)
75. Drossman DA, Corazziari E, Delvaux M, Spiller RC, Talley NJ, Thompson WG, Whitehead WE. *Rome. III: The Functional Gastrointestinal Disorders*. 3rd edn. McLean, VA: Degnon Associates, Inc.; 2006.
76. Parsons CL. Diagnosing chronic pelvic pain of bladder origin. *J Reprod Med* 2004;**49**(Suppl. 3):235–42.
77. Stead ML, Crocombe WD, Fallowfield LJ, Selby P, Perren TJ, Garry R, Brown JM. Sexual activity questionnaires in clinical trials: acceptability to patients with gynaecological disorders. *Br J Obstet Gynaecol* 1999;**106**:50–4. <https://doi.org/10.1111/j.1471-0528.1999.tb08084.x>
78. Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. *J Res Pers* 2007;**41**:203–12. <https://doi.org/10.1016/j.jrp.2006.02.001>
79. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 1995;**7**:524–32. <https://doi.org/10.1037/1040-3590.7.4.524>
80. Jones G, Jenkinson C, Kennedy S. Development of the Short Form Endometriosis Health Profile Questionnaire: the EHP-5. *Qual Life Res* 2004;**13**:695–704. <https://doi.org/10.1023/B:QURE.0000021321.48041.0e>
81. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003;**41**:1284–92. <https://doi.org/10.1097/01.MLR.0000093487.78664.3C>
82. Leserman J, Drossman DA, Li Z. The reliability and validity of a sexual and physical abuse history questionnaire in female patients with gastrointestinal disorders. *Behav Med* 1995;**21**:141–50. <https://doi.org/10.1080/08964289.1995.9933752>
83. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. [https://doi.org/10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9)

84. Al-Janabi H, Flynn T, Coast J. Development of a self-reported measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res* 2010;**21**:167–76.
85. The American Fertility Society. Classification of endometriosis. *Fertil Steril* 1979;**32**:633–4. [https://doi.org/10.1016/S0015-0282\(16\)44409-2](https://doi.org/10.1016/S0015-0282(16)44409-2)
86. Clopper C, Pearson E. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;**26**:404–13. <https://doi.org/10.1093/biomet/26.4.404>
87. Department of Health and Social Care (DHSC). *NHS Reference Costs 2013–14*. London: DHSC; 2014. URL: www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014 (accessed November 2014).
88. Curtis L. *Unit Costs of Health and Social Care 2014*. Canterbury: PSSRU, University of Kent; 2014.
89. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2008.
90. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal*. 2013. URL: www.nice.org.uk/process/pmg9 (accessed November 2014).

Appendix 1 Additional baseline data from the diagnostic study

TABLE 29 Recruitment by centre

Recruiting centre	n (%)
Basildon University Hospital	3 (1)
Birmingham Women's Hospital	6 (2)
Chelsea and Westminster	24 (8)
Cumberland Infirmary	6 (2)
Dorset County Hospital	3 (1)
Royal Infirmary of Edinburgh	5 (2)
Furness General Hospital	6 (2)
Homerton University Hospital	19 (7)
Liverpool Women's Hospital	11 (4)
Medway Maritime Hospital	2 (1)
Musgrove Park Hospital	9 (3)
Newham University Hospital	5 (2)
Ninewells Hospital	11 (4)
North Devon District Hospital	7 (2)
Nottingham University Hospital	4 (1)
Royal Hallamshire Hospital	14 (5)
The Royal London Hospital	80 (27)
Royal Preston Hospital	6 (2)
South Tyneside District Hospital	2 (1)
Southend University Hospital	7 (2)
Stafford Hospital	8 (3)
Sunderland Royal Hospital	6 (2)
University Hospital Crosshouse	10 (3)
University Hospital of North Durham	17 (6)
University Hospital of North Staffordshire	16 (6)
Whipps Cross University Hospital	4 (1)
Total	291 (100)

TABLE 30 Demographic details of participants

Demographic variable	Value
Height (cm), mean (SD, <i>n</i>)	1.64 (0.07, 283)
Weight (kg), mean (SD, <i>n</i>)	71.1 (15.3, 274)
Smoking	
Smoked > 100 cigarettes during lifetime, <i>n</i> (% , <i>N</i> responses)	117 (40.6, 288)
Current smoker, <i>n</i> (% , <i>N</i> responses)	67 (57.3, 117)
Units of alcohol drunk per week, mean (SD, <i>n</i>)	4.4 (6.5, 275)
Participant affected by special social circumstances, <i>n</i> (% , <i>N</i> responses)	17 (6.0, 283)

TABLE 31 Participant characteristics

Participant characteristics	Value
<i>Pain problems</i>	
<i>Duration of pain</i>	
Duration of pain (years), mean (SD, <i>n</i>)	4.2 (4.8, 287)
Event associated with the onset of pain, <i>n</i> (% , <i>N</i> responses)	136 (48.1, 283)
Change in pain since onset, <i>n</i> (%)	
Got a lot worse	155 (53.8)
Got a little worse	57 (19.8)
Not changed	59 (20.5)
Got a little better	12 (4.2)
Got a lot better	2 (0.7)
Do not know	3 (1.0)
Missing	3 (1.0)
Number of days with pelvic pain in previous month	
< 1 day per month	4 (1.4)
1 day per month	7 (2.5)
2–3 days per month	47 (16.6)
1 day per week	12 (4.2)
> 1 day per week	100 (35.2)
Every day	114 (40.1)
Missing	7 (2.4)
<i>Location of pain</i>	
Severity of pain in each area (0 = no pain, 10 = most severe pain imaginable), mean (SD, <i>n</i>)	
Left upper back quadrant	0.1 (0.9, 291)
Central upper back	0.2 (1.0, 291)
Right upper back quadrant	0.1 (0.7, 291)
Left lumbar back region	1.2 (2.8, 291)
Central back	1.5 (3.0, 291)
Right lumbar back region	1.2 (2.7, 291)

TABLE 31 Participant characteristics (*continued*)

Participant characteristics	Value
Left lower back region	1.0 (2.6, 291)
Central lower back	1.3 (3.0, 291)
Right lower back	1.1 (2.8, 291)
Left outer posterior thigh	0.3 (1.5, 291)
Left inner posterior thigh	0.5 (1.8, 291)
Right outer posterior thigh	0.3 (1.5, 291)
Right inner posterior thigh	0.4 (1.7, 291)
Right hypochondriac upper	0.1 (0.9, 291)
Epigastric region	0.1 (0.9, 291)
Left hypochondriac upper	0.1 (0.9, 291)
Right lumbar	1.0 (2.5, 291)
Umbilical region	1.3 (2.9, 291)
Left lumbar	1.0 (2.6, 291)
Right iliac	3.8 (4.1, 291)
Hypogastric/suprapubic	4.5 (4.1, 291)
Left iliac	3.7 (4.1, 291)
Right outer anterior thigh	0.5 (1.8, 291)
Right inner anterior thigh	0.5 (1.9, 291)
Left inner anterior thigh	0.5 (1.7, 291)
Left outer anterior thigh	0.4 (1.7, 291)
Urethral region	0.4 (1.7, 291)
Vulval	0.7 (2.2, 291)
Perianal	0.5 (1.9, 291)
Right inner thigh	0.2 (1.3, 291)
Right buttock	0.1 (0.7, 291)
Left inner thigh	0.2 (1.3, 291)
Left buttock	0.02 (0.4, 291)
Pain just before period (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	6.4 (2.6, 264)
Duration in months, mean (SD, <i>n</i>)	24.1 (22.5, 201)
Pain during period (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	7.1 (2.8, 266)
Duration in months, mean (SD, <i>n</i>)	25.3 (23.7, 202)
Pain when period is over (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	4.8 (2.8, 263)
Duration in months, mean (SD, <i>n</i>)	22.8 (22.0, 184)
Pain mid-cycle (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	5.4 (2.9, 263)
Duration in months, mean (SD, <i>n</i>)	23.7 (22.2, 193)
Sexual intercourse in the last month, <i>n</i> (%), <i>N</i> responses)	212 (73.4, 289)
If yes, pain at the point of vaginal penetration (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	3.6 (3.3, 209)
Duration in months, mean (SD, <i>n</i>)	20.3 (21.4, 114)
If yes, deep pain during intercourse (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	6.0 (3.3, 212)
Duration in months, mean (SD, <i>n</i>)	20.3 (20.7, 146)

continued

TABLE 31 Participant characteristics (*continued*)

Participant characteristics	Value
If yes, burning vaginal pain during intercourse (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	2.9 (3.4, 210)
Duration in months, mean (SD, <i>n</i>)	19.8 (18.5, 77)
If yes, pelvic pain lasting hours or days after (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	5.1 (3.3, 211)
Duration in months, mean (SD, <i>n</i>)	20.1 (21.1, 134)
Pain when bladder is full (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	4.1 (3.3, 286)
Duration in months, mean (SD, <i>n</i>)	18.5 (21.1, 146)
Pain with urination (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	2.1 (2.8, 283)
Duration in months, mean (SD, <i>n</i>)	17.1 (19.7, 95)
Muscle/joint pain in pelvic region (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	4.2 (3.6, 278)
Duration in months, mean (SD, <i>n</i>)	22.5 (22.7, 137)
Pain in pelvis when lifting (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	3.5 (3.4, 282)
Duration in months, mean (SD, <i>n</i>)	21.3 (22.7, 133)
Pain when sitting (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	4.0 (3.3, 282)
Duration in months, mean (SD, <i>n</i>)	19.2 (18.1, 147)
Backache (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	5.1 (3.3, 286)
Duration in months, mean (SD, <i>n</i>)	22.7 (24.5, 168)
Migraine headache (0 = no pain, 10 = worst pain imaginable), mean (SD, <i>n</i>)	3.4 (3.7, 222)
Duration in months, mean (SD, <i>n</i>)	23.8 (23.5, 93)
<i>Treatments for pain, n (%)</i>	
Acupuncture	13 (4.5)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	7 (53.8, 13)
Anti-seizure medications	7 (2.4)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	2 (28.6, 7)
Antidepressants	39 (13.4)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	17 (47.2, 36)
Biofeedback	1 (0.3)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	1 (100, 1)
Botox injection	2 (0.7)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	1 (50, 2)
Contraceptive pills/patch/ring	162 (55.7)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	51 (32.7, 156)
Exercise, yoga or pilates	131 (45.0)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	38 (30.4, 125)
Hormonal therapy for endometriosis	27 (9.3)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	10 (40, 25)
Herbal medicine	27 (9.3)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	8 (32, 25)
Homeopathic medicine	14 (4.8)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	3 (25, 12)

TABLE 31 Participant characteristics (*continued*)

Participant characteristics	Value
Massage	80 (27.5)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	32 (42.7, 75)
Meditation or relaxation exercises	57 (19.6)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	22 (40, 55)
Strong painkillers	237 (81.4)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	137 (60.9, 225)
Nerve blocks	5 (1.7)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	2 (40, 5)
Non-prescription medicine	68 (23.4)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	26 (41.3, 63)
Nutrition/diet	90 (30.9)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	28 (33.7, 83)
Physiotherapy	27 (9.3)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	5 (19.2, 26)
Psychological (talking) therapy	20 (6.9)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	6 (33.3, 18)
TENS	12 (4.1)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	6 (55.5, 11)
Other	28 (9.6)
If used, was considered helpful, <i>n</i> (% , <i>N</i> responses)	15 (57.7, 26)
Menstrual history	
Age at menarche, mean (SD, <i>n</i>)	12.8 (1.7, 290)
Currently having menstrual periods, <i>n</i> (% , <i>N</i> responses)	252 (86.9, 290)
Pelvic pain with periods in at 3 months, <i>n</i> (% , <i>N</i> responses)	
No	13 (5.2)
Occasionally (with 1 in 3 of my periods)	19 (7.6)
Often (with 2 in 3 of my periods)	32 (12.9)
Always (every period)	185 (74.3)
Missing	42
Pelvic pain at times other than with periods or sexual intercourse in the last 3 months, <i>n</i> (% , <i>N</i> responses)	223 (90.6, 246)
If yes, pain before period, <i>n</i> (% , <i>N</i> responses)	197 (88.3, 223)
If yes, pain after period, <i>n</i> (% , <i>N</i> responses)	145 (66.1, 221)
Period regularity, <i>n</i> (%)	
Regular, I know when to expect my period	89 (35.9)
Fairly regular, my period starts within a few days of when I expect	75 (30.2)
Irregular, I cannot predict when my period will start	53 (21.4)
I have bleeding on and off all the time	31 (12.5)
Missing	43

continued

TABLE 31 Participant characteristics (continued)

Participant characteristics	Value
Period heaviness, <i>n</i> (%)	
Light	28 (11.2)
Moderate	76 (30.3)
Heavy	95 (37.9)
Bleed through protection	52 (20.7)
Missing	40
Duration of period (days), mean (SD, <i>n</i>)	5.3 (2.0, 244)
Average interval between periods (days), mean (SD, <i>n</i>)	28.0 (12.3, 211)
Clots in menstrual flow, <i>n</i> (%), <i>N</i> responses	198 (80.5, 246)
Pain on day flow starts, <i>n</i> (%), <i>N</i> responses	53 (22.2, 239)
If no, how many days before flow, mean (SD, <i>n</i>)	4.3 (3.3, 152)
Previous diagnoses, <i>n</i> (%)	
Endometriosis	76 (26.7)
Adhesions	20 (7.0)
Fibroids	35 (12.3)
Adenomyosis	10 (3.5)
Uterine polyps	10 (3.5)
Ovarian cysts	105 (36.8)
Appendicitis	29 (10.2)
Hernia	10 (3.5)
Infertility or low fertility	25 (8.8)
Uterine or bladder prolapse	2 (0.7)
Vulva pain/vulvodynia	8 (2.8)
Irritable bowel syndrome	49 (17.2)
Nerve entrapment in pelvis/pudendal neuropathy	1 (0.4)
Fibromyalgia	7 (2.5)
Painful bladder syndrome (interstitial cystitis)	12 (4.2)
Sexually transmitted infection	36 (12.7)
Female genital mutilation/cutting	2 (0.7)
<i>Previous tests</i>	
Previous cervical screening test, <i>n</i> (%), <i>N</i> responses	229 (79, 290)
If test performed, outcome, <i>n</i> (%)	
Normal	210 (92.5)
Abnormal	17 (7.5)
Missing	2 (0.9)
Previous chlamydia test, <i>n</i> (%), <i>N</i> responses	202 (70.4, 287)
If test performed outcome, <i>n</i> (%)	
Negative for chlamydia	176 (87.1)
Positive for chlamydia and treated	21 (10.4)
Positive for chlamydia, not treated	0 (–)
Missing	5 (2.4)

TABLE 31 Participant characteristics (*continued*)

Participant characteristics	Value
<i>Previous investigations/operations, n (%)</i>	
Previous investigations/operations	
Laparoscopy	80 (27.5)
Cystoscopy	16 (5.5)
Colonoscopy	36 (12.4)
Hysteroscopy	24 (8.3)
MRI	30 (10.3)
Laparotomy	9 (3.1)
Transvaginal ultrasound	199 (68.4)
Transabdominal ultrasound	220 (75.6)
Nerve transmission test	1 (0.3)
Allergy tests	20 (6.9)
Other	7 (2.4)
Contraceptive user, <i>n (%)</i>	172 (71.4)
Of those using contraception, method(s) used	
Female sterilisation (clips)	9 (5.2)
Female sterilisation (implants)	1 (0.6)
Male partner sterilisation	11 (6.4)
Contraceptive pill	71 (41.3)
Mini pill	12 (7.0)
Injection	3 (1.7)
Patch	1 (0.6)
Implant	10 (5.8)
Levonorgestrel-releasing intrauterine system	19 (11.1)
Condom	67 (39.0)
Diaphragm/cap	1 (0.6)
Vaginal ring	–
Natural method	7 (4.1)
Obstetric history	
<i>Fertility, n (%)</i>	
Currently trying to get pregnant, <i>n (%)</i> , <i>N</i> responses	36 (14.8, 242)
Of those trying to get pregnant, <i>n (%)</i>	
Trying for < 1 year	8 (22.2)
Trying for > 1 year	28 (77.8)
<i>Pregnancies, n (%)</i>	
Number of times been pregnant previously	
0	113 (40.0)
1	46 (16.0)
2	51 (18.0)
3	33 (12.0)
continued	

TABLE 31 Participant characteristics (continued)

Participant characteristics	Value
4	20 (7.0)
≥ 5	19 (7.0)
Missing	9
If at least one previous pregnancy, outcome	
Live birth	
0	38 (22.4)
1	41 (24.2)
2	58 (34.3)
3	23 (13.6)
4	5 (2.9)
≥ 5	1 (0.6)
Missing	3 (1.8)
Termination	
0	110 (65.1)
1	47 (27.8)
2	8 (4.7)
3	3 (1.8)
4	–
≥ 5	–
Missing	1 (0.6)
Stillbirth	
0	158 (93.5)
1	7 (4.1)
2	1 (0.6)
3	–
4	1 (0.6)
≥ 5	–
Missing	2 (1.2)
Miscarriage	
0	112 (66.3)
1	35 (20.7)
2	10 (5.9)
3	3 (1.8)
4	4 (2.4)
≥ 5	3 (1.7)
Missing	2 (1.2)
If at least one live birth, route of delivery	
Vaginal	
0	79 (47.6)
1	27 (16.3)

TABLE 31 Participant characteristics (*continued*)

Participant characteristics	Value
2	41 (24.7)
3	14 (8.4)
4	4 (2.4)
≥ 5	1 (0.6)
Caesarean section (planned)	
0	151 (91.0)
1	10 (6.0)
2	4 (2.4)
3	1 (0.6)
4	–
≥ 5	–
Caesarean section (emergency)	
0	135 (80.8)
1	26 (15.6)
2	5 (3.0)
3	–
4	1 (0.6)
≥ 5	–
Forceps/ventouse	
0	146 (88.0)
1	18 (10.8)
2	2 (1.2)
3	–
4	–
≥ 5	–
Did any of the following occur during pregnancy, labour or just after giving birth (yes)	
Abdominal pain (other than labour contractions)	56 (37)
Perineal tear	53 (35)
Episiotomy	37 (25)
Pelvic girdle pain	26 (17)
Interventions for peri-partum haemorrhage	23 (15)
Participant-reported questionnaires	
<i>Pelvic Pain and Urgency/Frequency Patient Symptom Scale</i>	
Summary score (points) (SD, n)	13.3 (5.5, 167)
<i>Assessment of sexual activity</i>	
Section I, n (%)	
Are you currently married or having an intimate relationship with someone?	238 (83)
Have you changed your sexual partner in the last 6 months?	24 (8)
Do you engage in sexual activity with anyone at the moment?	222 (78)
continued	

TABLE 31 Participant characteristics (*continued*)

Participant characteristics	Value
Section II, n (%)	
I am not sexually active at the moment because	
I do not have a partner at the moment	37 (60)
I am too tired	6 (10)
My partner is too tired	1 (2)
I am not interested in sex	10 (16)
My partner is not interested in sex	3 (5)
I have a physical problem that makes sexual relations difficult or uncomfortable	13 (22)
My partner has a physical problem that makes sexual relations difficult or uncomfortable	1 (2)
Other reason	13 (22)
Section III	
Pleasure score (0 = worst outcome, 18 = best outcome) (SD, n)	7.4 (4.5, 104)
Discomfort score (0 = worst outcome, 6 = best outcome) (SD, n)	3.1 (1.7, 204)
Habit score (0 = less than usual, 1 = the same as usual, 2 = somewhat more than usual, 3 = much more than usual) (SD, n)	2.2 (0.8, 213)
BFI score	
Extraversion (0 = worst outcome, 10 = best outcome), mean (SD, n)	7.1 (2.0, 283)
Agreeableness (0 = worst outcome, 10 = best outcome), mean (SD, n)	7.6 (1.8, 284)
Conscientiousness (0 = worst outcome, 10 = best outcome), mean (SD, n)	8.4 (1.6, 279)
Neuroticism (0 = worst outcome, 10 = best outcome), mean (SD, n)	6.2 (2.1, 286)
Openness (0 = worst outcome, 10 = best outcome), mean (SD, n)	6.9 (1.6, 282)
Pain Catastrophizing Scale	
Over the last month, how much have you been affected by your pain (0 = least affected, 20 = worst affected), mean (SD, n)	–
Of all the problems and stresses in your life, how does your pain compare in importance? (0 = just one of many problems, 10 = the most important thing), mean (SD, n)	–
Depression (6 = worst outcome, 0 = best outcome), mean (SD, n)	2.2 (1.8, 289)
Rumination (16 = worst outcome, 0 = best outcome), mean (SD, n)	8.5 (4.8, 285)
Magnification (12 = worst outcome, 0 = best outcome), mean (SD, n)	5.4 (3.4, 284)
Helplessness (24 = worst outcome, 0 = best outcome), mean (SD, n)	11.6 (6.4, 284)
EQ-5D-3L	
Summary score (–0.59 = worst outcome, 1.0 = best outcome) (SD, n)	0.6 (0.3, 286)
Health state (0 = worst outcome, 100 = best outcome) (SD, n)	63.4 (19.4, 270)
ICECAP	
Capabilities, as a measure of well-being (0 = worst outcome, 1.0 = best outcome) (SD, n)	0.3 (0.2, 198)
BFI, Big Five Inventory; TENS, transcutaneous electrical nerve stimulation.	

Appendix 2 Additional results from the diagnostic study

TABLE 32 Accuracy estimates for establishing the absence of any gynaecological cause against laparoscopy and the EIP assessment without MRI findings

	Reference standard								
	Laparoscopy assessment of absence of any gynaecological cause			EIP A assessment of absence of any gynaecological cause			EIP B assessment of absence of any gynaecological cause		
Features of absence of any gynaecological cause in the MRI report		Yes	No		Yes	No		Yes	No
	Yes	16	33	Yes	26	23	Yes	27	22
	No	59	179	No	130	108	No	126	112
<i>p</i> -value (Fisher's exact test)	<i>p</i> = 0.3			<i>p</i> = 0.9			<i>p</i> = 0.9		
Accuracy estimates (95% CI)	Sens = 21.3 (12.7 to 32.3)			Sens = 16.7 (11.2 to 23.5)			Sens = 17.7 (12.0 to 24.6)		
	Spec = 84.4 (78.8 to 89.0)			Spec = 82.4 (74.8 to 88.5)			Spec = 83.6 (76.2 to 89.4)		
	LR+ = 1.4 (0.8 to 2.3)			LR+ = 0.9 (0.6 to 1.6)			LR+ = 1.1 (0.6 to 1.8)		
	LR− = 0.9 (0.8 to 1.1)			LR− = 1.0 (0.9 to 1.1)			LR− = 1.0 (0.9 to 1.1)		
	DOR = 1.5 (0.8 to 2.9)			DOR = 0.9 (0.5 to 1.7)			DOR = 1.1 (0.6 to 2.0)		
DOR, diagnostic odds ratio; LR+, likelihood ratio (positive); LR−, likelihood ratio (negative); Sens, sensitivity; Spec, specificity.									

TABLE 33 Accuracy estimates for establishing the absence of any gynaecological cause against laparoscopy and EIP with MRI findings

	Reference standard								
	Laparoscopy assessment of absence of any gynaecological cause			EIP A assessment of absence of any gynaecological cause			EIP B assessment of absence of any gynaecological cause		
Features of absence of any gynaecological cause in the MRI report		Yes	No		Yes	No		Yes	No
	Yes	55	93	Yes	88	60	Yes	88	60
	No	20	119	No	68	71	No	65	74
<i>p</i> -value (Fisher's exact test)	<i>p</i> < 0.0001			<i>p</i> = 0.08			<i>p</i> = 0.03		
Accuracy estimates (95% CI)	Sens = 73.3 (61.9 to 82.9)			Sens = 56.4 (48.3 to 64.3)			Sens = 57.5 (49.3 to 65.5)		
	Spec = 56.1 (49.2 to 62.9)			Spec = 54.2 (45.3 to 62.9)			Spec = 55.2 (46.4 to 63.8)		
	LR+ = 4.7 (3.3 to 6.6)			LR+ = 1.2 (0.9 to 1.5)			LR+ = 0.4 (0.3 to 0.6)		
	LR− = 0.3 (0.2 to 0.5)			LR− = 0.8 (0.6 to 1.0)			LR− = 1.5 (1.3 to 1.8)		
	DOR = 14.9 (8.4 to 26.6)			DOR = 1.5 (0.9 to 2.4)			DOR = 0.3 (0.2 to 0.4)		
DOR, diagnostic odds ratio; LR+, likelihood ratio (positive); LR−, likelihood ratio (negative); Sens, sensitivity; Spec, specificity.									

TABLE 34 Accuracy of independent gynaecologist's diagnosis before and after MRI

Confidence score	Pre-index data diagnosis			Pre-index data and MRI report diagnosis		
	EIP A judgement: absence of any gynaecological cause of pain, <i>n</i> (%)		Likelihood ratio (95% CI)	EIP A judgement: absence of any gynaecological cause of pain, <i>n</i> (%)		Likelihood ratio (95% CI)
	Yes	No		Yes	No	
7–10	53 (34.0)	43 (32.8)	1.0 (0.7 to 1.4)	77 (49.4)	44 (33.6)	1.5 (1.1 to 2.0)
4–6	86 (55.1)	69 (52.7)	1.0 (0.8 to 1.3)	49 (31.4)	47 (35.9)	0.9 (0.6 to 1.2)
0–3	17 (10.9)	19 (14.5)	0.8 (0.4 to 1.4)	30 (19.2)	40 (30.5)	0.6 (0.4 to 1.0)
Total	156	131	$p^* = 0.5$	156	131	$p^* = 0.004$

p^* , p -value computed from a chi-squared test for trend.

TABLE 35 Accuracy of treating gynaecologist's diagnosis before and after laparoscopy

Confidence score	Pre-index data diagnosis			Pre-index data and laparoscopy report		
	EIP A judgement: absence of any gynaecological cause of pain, <i>n</i> (%)		Likelihood ratio (95% CI)	EIP A judgement: absence of any gynaecological cause of pain, <i>n</i> (%)		Likelihood ratio (95% CI)
	Yes	No		Yes	No	
7–10	62 (40.0)	37 (28.2)	1.5 (1.1 to 2.1)	81 (53.6)	4 (3.1)	17.6 (6.6 to 46.6)
4–6	60 (38.7)	34 (26.0)	1.5 (1.1 to 2.1)	26 (17.2)	18 (13.7)	1.3 (0.7 to 2.2)
0–3	33 (21.3)	60 (45.8)	0.5 (0.3 to 0.7)	44 (29.1)	109 (83.2)	0.4 (0.3 to 0.5)
Total	155	131	$p^* = 0.0002$	151	131	$p^* < 0.0001$

p^* , p -value computed from a chi-squared test for trend.

Appendix 3 Economic evaluation sensitivity analyses

The distributions used in the PSA for the economic model are shown in *Table 36*.

TABLE 36 Distributions used in the PSA for all scenarios

Parameter	Distribution	Alpha	Beta
% of eligible women who receive a laparoscopy when offered	Beta	287	20
% of eligible women who receive a MRI when offered	Beta	287	35
% of women who have a therapeutic laparoscopy during the same procedure as a diagnostic laparoscopy	Beta	92	20
% of women who receive a second therapeutic laparoscopy during a separate procedure following their first therapeutic laparoscopy	Beta	30	82
Prevalence (negative, positive for therapeutic laparoscopy, other treatment)	Dirichlet (164,112,11)		
Proportion of women who require other treatment who have fibroids (remainder of women have adenomyosis)	Beta	3	8
Time for consultant and nurse to conduct baseline examination	Gamma	1	0.33

A decorative graphic consisting of numerous thin, parallel green lines that curve from the left side of the page towards the right, creating a sense of movement and flow.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library